NOTE

This Microfiche meets the quality requirements of the American National Standard for Microfiche of Documents (ANSI PH5.9-1975). However, because of imperfections in the document from which it was made, it contains images that are illegible or difficult to read. The source document for this microfiche is the best copy that could be obtained prior to filming.

CIS Congressional Information Service,
Washington, D.C.

S. Hrg. 101-638

NATIONAL HIGH-PERFORMANCE COMPUTER TECHNOLOGY ACT OF 1989

HEARINGS

BEFORE THE

SUBCOMMITTEE ON SCIENCE, TECHNOLOGY, AND SPACE

OF THE

COMMITTEE ON COMMERCE, SCIENCE, AND TRANSPORTATION
UNITED STATES SENATE

ONE HUNDRED FIRST CONGRESS

FIRST SESSION

ON

S. 1067

TO PROVIDE FOR A COORDINATED FEDERAL RESEARCH PROGRAM TO ENSURE CONTINUED UNITED STATES LEADERSHIP IN HIGH-PERFORMANCE COMPUTING

JUNE 21, JULY 26, AND SEPTEMBER 15, 1989

Printed for the use of the
Committee on Commerce, Science, and Transportation

U.S. GOVERNMENT PRINTING OFFICE
27-517 O WASHINGTON : 1989

S 261-50

For sale by the Superintendent of Documents, Congressional Sales Office
U.S. Government Printing Office, Washington, DC 20402

COMMITTEE ON COMMERCE, SCIENCE, AND TRANSPORTATION

ERNEST F. HOLLINGS, South Carolina, *Chairman*

DANIEL K. INOUYE, Hawaii
WENDELL H. FORD, Kentucky
J. JAMES EXON, Nebraska
ALBERT GORE, JR., Tennessee
JOHN D. ROCKEFELLER IV, West Virginia
LLOYD BENTSEN, Texas
JOHN F. KERRY, Massachusetts
JOHN B. BREAUX, Louisiana
RICHARD H. BRYAN, Nevada
CHARLES S. ROBB, Virginia

JOHN C. DANFORTH, Missouri
BOB PACKWOOD, Oregon
LARRY PRESSLER, South Dakota
TED STEVENS, Alaska
ROBERT W. KASTEN, JR., Wisconsin
JOHN McCAIN, Arizona
CONRAD BURNS, Montana
SLADE GORTON, Washington
TRENT LOTT, Mississippi

RALPH B. EVERETT, *Chief Counsel and Staff Director*
WALTER B. MCCORMICK, JR., *Minority Chief Counsel and Staff Director*

SUBCOMMITTEE ON SCIENCE, TECHNOLOGY, AND SPACE

ALBERT GORE, JR., Tennessee, *Chairman*

JOHN D. ROCKEFELLER IV, West Virginia
LLOYD BENTSEN, Texas
JOHN F. KERRY, Massachusetts
RICHARD H. BRYAN, Nevada
CHARLES S. ROBB, Virginia

LARRY PRESSLER, South Dakota
TED STEVENS, Alaska
ROBERT W. KASTEN, JR., Wisconsin
TRENT LOTT, Mississippi

CONTENTS

Page
Opening statement by Senator Gore 1
Text of S. 1067 3

CHRONOLOGICAL LIST OF WITNESSES

JUNE 21, 1989

Collins, A. Gray, Jr., senior vice president, external affairs, Bell Atlantic Corp 103
Fields, Dr. Craig, director, Defense Advanced Research Projects Agency 57
  Prepared statement 63
Gabbard, O. Gene, chairman and CEO, Telecom*USA 104
  Prepared statement 108
Handler, Sheryl L., president, Thinking Machines Corp 39
Liebhaber, Richard T., executive vice president, MCI Communications Corp 95
  Prepared statement 97
Lucky, Robert W., executive director, research, communications sciences, AT&T Bell Laboratories 82
  Prepared statement 84
Rollwagen, John A., chairman and CEO, Cray Research, Inc 24
  Prepared statement 30
Schwantes, Roger, vice president for product and technology planning, Northern Telecom, Inc 120
Wulf, Dr. William, Assistant Director, Computer and Information Science and Engineering, NSF 65
  Prepared statement 69

JULY 26, 1989

Opening statement by Senator Gore 127
Opening statement by the Chairman 130
Opening statement by Senator Pressler 160
Clark, Dr. James H., chairman and founder, Silicon Graphics 185
  Prepared statement 190
Nagel, Dr. David, manager of user technologies, advanced technology group, Apple Computer, Inc 205
  Prepared statement 209
Poduska, Dr. J. William, Sr., chairman and CEO, Stellar Computer, Inc.; accompanied by Norm Siegel, systems engineer 196
  Prepared statement 199
Reddy, Dr. Ray, director, robotics institute, Carnegie-Mellon University 223
  Prepared statement 228
Schwartz, Dr. Jacob T., director, information science and technology office, Defense Advanced Research Projects Agency 141
  Prepared statement 145
Winkler, Dr. Karl-Heinz A., deputy director for science technololgy, and education, national center for supercomputing applications, University of Illinois 168
  Prepared statement 175
Wulf, Dr. William, Assistant Director, Computer and Information Science and Engineering, NSF 131
  Prepared statement 134

SEPTEMBER 15, 1989

Opening statement by Senator Gore 243
Opening statement by Senator Pressler 300
Billington, Hon. James H., Librarian, Library of Congress; accompanied by Henriette Avram, Assistant Librarian for Processing Services 259
  Prepared statement 264
  Questions of Senator Gore and the answers 270
Brown, John Seely, vice president, advanced research, Xerox Palo Alto Research Center 347
  Prepared statement 351
Fischer, Dr. John N., Acting Associate Director, USGS; accompanied by David A. Nystrom, Chief, Office of Geographic and Cartographic Reasearch; and Lawrence Batten 308
  Prepared statement 311
  Questions of Senator Gore and the answers 317
Kahn, Robert E., president, Corporation for National Research Initiatives 277
  Prepared statement 282
Masys, Dr. Daniel S., director, Lister Hill National Center For Biomedical Communications, National Library of Medicine 330
  Prepared statement 334
  Questions of Senator Gore and the answers 341
Nelson, Ted, Autodesk, Inc 398
  Prepared statement 401
Wladawaky, Dr. Irving, DSD vice president, IBM Corp 362
  Prepared statement 365
  Questions of Senator Gore and the answers 373
Wood, Richard T., senior vice president for business development, University Microfilms, Inc 387
  Prepared statement 389
  Questions of Senator Gore and the answers 393
Wyatt, Joe, chancellor, Vanderbilt University 254

ADDITIONAL ARTICLES, LETTERS, AND STATEMENTS

Association of American Publishers, statement 414
Kerry, Hon, Robert, U.S. Senator from Nebraska, statement 409

NATIONAL HIGH-PERFORMANCE COMPUTER TECHNOLOGY ACT OF 1989

WEDNESDAY, JUNE 21, 1989

U.S. Senate,
Committee on Commerce, Science, and Transportation,
Subcommittee on Science, Technology, and Space,
Washington, DC.

The subcommittee met, pursuant to notice, at 2 p.m., in room SR-253, Russell Senate Office Building, Hon. Albert Gore, Jr. (chairman of the subcommittee) presiding.

Staff members assigned to these hearings: Mike Nelson, professional staff member and Fiona Branton, minority staff counsel.

OPENING STATEMENT BY SENATOR GORE

Senator Gore. The Subcommittee will come to order. I want to apologize to our witnesses, our guests and my colleagues for being a few minutes late. I was held up unavoidably at the Capitol and I do apologize.

But I have really been looking forward to this hearing and to other discussions later on today and this evening. I am excited about the possibilities we will be discussing here and I look forward to the testimony very much.

We are going to examine today the enormous benefits supercomputing and computer networking can provide to American researchers, to American students, to American business, to the whole country. We will hear about how these new technologies can also greatly improve the U.S. competitiveness and about the Federal Government's proper role in promoting them.

Supercomputers will be in the 1990s what machine tools were in the 1970s and 1980s. I also like the analogy that has become a cliche, that supercomputers will be the steam engines of the Information Age.

During the last 20 years, new improved computerized machine tools have enabled workers to produce more and more quickly with less waste throughout the manufacturing sector, so although the machine tool industry represents a very small fraction of GNP, it is critical to the country's competitiveness, since its products impact hundreds of other industries and improve productivity throughout the economy.

In a similar way, with even more powerful effects, new supercomputers are improving productivity and profits in many American companies. Mechanical engineers use supercomputer models to design automobile parts that are more reliable, durable and easier to make.

Supercomputers are improving manufacturing on the shop floor and the quality of the car in your garage. Aeronautical engineers are using supercomputers instead of wind tunnels to test new aircraft designs, and this is often ten times cheaper and ten times faster than using a conventional wind tunnel.

Also, as this Committee knows, supercomputer models will be essential for developing NASA's national aerospace plane, because no wind tunnel on earth operates at the speeds at which NASP will fly.

There are many other uses that I will discuss in my prepared statement and I am not going to deliver all of this, particularly since we have got a little bit of a late start.

On May 18th, I introduced S. 1067, the National High Performance Computer Technology Act. It is a comprehensive bill aimed at enhancing the development and use of all aspects of advanced computing. It includes titles on computer hardware, software, networking and basic research and education.

Now, I do not think we can focus just on the hardware and our witnesses will tell us why it is important to look at all of the other aspects of computing, like software and education, and we do need a balanced approach.

Another analogy that I frequently draw is between the network we are talking about here and the interstate highway system. We need an interstate highway system of information. Superhighways. This bill will make it possible.

We are going to have hearings on the other titles of the bill next month, but today we are going to focus on Title 2, the title that deals with computer networking.

The bill calls for the National Science Foundation to establish by 1996 a National Research and Education Network capable of transmitting 3 billion bits of data every second. That is equivalent to 100,000 single space typewritten pages every second. That is the Encyclopaedia Britannica in a single second.

That sounds like a lot of information, and it is, but in the future scientists will have to handle even larger flows of data. We are now on the sixth or seventh generation of supercomputers, but we are only on the second generation of networks and we need to pay careful attention to the benefits the country can gain from moving forward rapidly toward the kind of information superhighway network that the bill envisions.

I want to just serve notice that this is going to be an extremely high priority for our Subcommittee and I am pleased at all of the support that has emerged for it.

Let me recognize my colleagues for any statements they might want to make. Senator Bryan?

Senator BRYAN. No, thank you.

[The bill follows:]

101ST CONGRESS
1ST SESSION

S. 1067

To provide for a coordinated Federal research program to ensure continued United States leadership in high-performance computing.

IN THE SENATE OF THE UNITED STATES

MAY 18 (legislative day, JANUARY 3), 1989

Mr. GORE (for himself and Mr. JEFFORDS) introduced the following bill; which was read twice and referred to the Committee on Commerce, Science, and Transportation

---

A BILL

To provide for a coordinated Federal research program to ensure continued United States leadership in high-performance computing.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. This Act may be cited as the "National High-Performance Computer Technology Act of 1989".

SEC. 2. (a) Congress finds and declares the following:

(1) Advances in computer science and technology are vital to the Nation's prosperity, national security, and scientific advancement.

(2) The United States currently leads the world in development and use of high-performance computer technology for national security, industrial productivity, and science and engineering, but that lead is being challenged by foreign competitors.

(3) Further research and improved computer research networks are necessary to maintain United States leadership in the field of high-performance computing.

(b) It is the purpose of Congress in this Act to ensure the continued leadership of the United States in high-performance computer technology. This requires that the United States Government—

(1) expand Federal support for research, development, and application of high-performance computing technology in order to—

(A) establish a high-capacity national research and education computer network;

(B) develop an information infrastructure of data bases, services, and knowledge banks which is available for access over such a national network;

(C) promote the more rapid development and wider distribution of computer software;

(D) stimulate research on artificial intelligence;

(E) accelerate the development of computer systems; and

(F) invest in basic research and education; and

(2) improve planning and coordination of Federal research and development on high-performance computing.

TITLE I—NATIONAL HIGH-PERFORMANCE COMPUTER TECHNOLOGY PROGRAM

SEC. 101. The National Science and Technology Policy, Organization, and Priorities Act of 1976 (42 U.S.C. 6601 et seq.) is amended by adding at the end the following new title:

"TITLE VI—NATIONAL HIGH-PERFORMANCE COMPUTER TECHNOLOGY PROGRAM

"FINDINGS

"SEC. 601. (a) Congress finds and declares the following:

"(1) In order to strengthen America's computer industry and to assist the entire manufacturing sector, the Federal Government must provide leadership in the development and application of high-performance computer technology. In particular, the Federal Government should support the development of a high-capacity, national research and education network; facilitate the development of software for research, education, and industrial applications; continue to fund basic research; and provide for the training of computer scientists and computational scientists.

"(2) Several Federal agencies have ongoing high-performance computer technology programs. Improved interagency coordination, cooperation, and planning could enhance the effectiveness of these programs.

"(3) A recent report by the Office of Science and Technology Policy outlining a research and development strategy for high-performance computing provides a framework for a multiagency computer technology program.

"NATIONAL HIGH-PERFORMANCE COMPUTER TECHNOLOGY PLAN

"Sec. 602.(a)(1) The President, through the Federal Coordinating Council for Science, Engineering, and Technology (hereafter in this title referred to as the 'Council'), shall develop and implement a National High-Performance Computer Technology Plan (hereafter in this title referred to as the 'Plan') in accordance with the provisions, findings, and purpose of this Act. Consistent with the responsibilities set forth under subsection (c) of this section, the Plan shall contain recommendations for a five-year national effort, to be submitted to Congress within one year after the date of enactment of this title and to be revised at least once every two years thereafter.

"(2) The Plan shall—

"(A) establish the goals and priorities for a Federal high-performance computer technology program for the fiscal year in which the Plan (or revised Plan) is submitted and the succeeding four fiscal years;

"(B) set forth the role of each Federal agency and department in implementing the Plan;

"(C) describe the levels of Federal funding and specific activities, including education, research activities, hardware and software development, and acquisition and operating expenses for computers and computer networks, required to achieve such goals and priorities; and

"(D) consider and use, as appropriate, reports and studies conducted by Federal agencies and departments, the National Research Council, or other entities.

"(3) The Plan shall address, where appropriate, the relevant programs and activities of the following Federal agencies and departments—

"(A) the National Science Foundation;

"(B) the Department of Commerce, particularly the National Institute of Standards and Technology and the National Oceanic and Atmospheric Administration;

"(C) the National Aeronautics and Space Administration;

"(D) the Department of Defense, particularly the Defense Advanced Research Projects Agency, the Office of Naval Research, and, as appropriate, the National Security Agency;

"(E) the Department of Energy;

"(F) the Department of Health and Human Services, particularly the National Institutes of Health; and

"(G) such other agencies and departments as the President or the Chairman of the Council considers appropriate.

"(b) The Council shall—

"(1) serve as lead entity responsible for development and implementation of the Plan;

"(2) coordinate the high-performance computing research and development activities of Federal agencies and departments and report at least annually to the President, through the Chairman of the Council, on any recommended changes in agency or departmental roles that are needed to better implement the Plan;

"(3) prior to the President's submission to Congress of the annual budget estimate, review each agency and departmental budget estimate in the context of the Plan and make the results of that review available to each agency and department and to the appropriate elements of the Executive Office of the President, particularly the Office of Management and Budget;

"(4) work with Federal agencies, with the National Research Council and with academic, State, and other groups conducting research on high-performance computing; and

"(5) consult with actual and potential users of such research by establishing an advisory board, which shall include representatives from universities and industry.

"(c)(1) The Plan shall take into consideration, but not be limited to, the following missions and responsibilities of agencies and departments:

"(A) The National Science Foundation shall continue to be responsible for basic research in all areas of computer science, materials science, and computational science. The Foundation shall continue to solicit grant proposals and award grants by merit review for research in universities, nonprofit research institutions, and industry. The National Science Foundation shall also be responsible for providing researchers with access to supercomputers and providing for the establishment, by 1996, of a three-gigabit-per-second national computer network, as required by section 201 of the National High-Performance Computer Technology Act of 1989. Additional responsibilities include development of an information infrastructure of services, data bases, and knowledge banks connected to such computer network; facilitation of the validation of software and distribution of that software over such computer network; and promotion of science and engineering education.

"(B) The National Institute of Standards and Technology shall be responsible for ensuring interoperability between computer networks run by different agencies of the Federal Government and for establishing, in conjunction with industry, benchmark tests and standards for high-performance computers and software. Pursuant to the Computer Security Act of 1987 (Public Law 100-235; 100 Stat. 1724), the National Institute of Standards and Technology shall continue to be responsible for developing standards and guidelines for Federal computer systems, including standards and guidelines needed to assure the cost-effective security and privacy of sensitive information in Federal computer systems.

"(C) The National Oceanic and Atmospheric Administration shall continue to observe, collect, communicate, analyze, process, provide, and disseminate data about the Earth, its oceans, atmosphere, and space environment. It shall improve the quality and accessibility of the environmental data stored at the four National Oceanic and Atmospheric Administration data centers. In addition, the National Oceanic and Atmospheric Administration shall perform research and develop technology to support its data handling role.

"(D) The National Aeronautics and Space Administration shall continue to conduct basic and applied research in high-performance computing, particularly in the field of computational science, with emphasis on aeronautical applications and remote sensing data processing.

"(E) The Department of Defense, through the Defense Advanced Research Projects Agency, the Office of Naval Research, and other agencies, shall continue to conduct basic and applied research in high-performance computing, particularly in computer networking, semiconductor technology, and large-scale parallel processors. Pursuant to the Stevenson-Wydler Technology Innovation Act of 1980 (15 U.S.C. 3701 et seq.), the Department shall ensure that unclassified computer technology research is readily available to American industry. The National Security Agency, pursuant to the Computer Security Act of 1987 (Public Law 100–235; 100 Stat. 1724), shall continue to provide, where appropriate, technical advice and assistance to the National Institute of Standards and Technology for the development of standards and guidelines needed to assure the cost-effective security and privacy of sensitive information in Federal computer systems.

"(F) The Department of Energy and its national laboratories shall continue to conduct basic and applied research in high-performance computing, particularly in software development and multiprocessor supercomputers. Pursuant to the Stevenson-Wydler Technology Innovation Act of 1980 (15 U.S.C. 3701 et seq.), and other appropriate statutes, the Department of Energy shall ensure that unclassified computer technology research is readily available to American industry.

"(2) The Plan shall facilitate collaboration among agencies and departments with respect to—

"(A) ensuring interoperability among computer networks run by the agencies and departments;

"(B) increasing software productivity, capability, and reliability;

"(C) promoting interoperability of software;

"(D) distributing software among the agencies and departments; and

"(E) distributing federally funded, unclassified software to industry and universities.

"(d)(1) Each Federal agency and department involved in high-performance computing shall, as part of its annual request for appropriations to the Office of Management and Budget, submit a report identifying each element of its high-performance computing activities, which—

"(A) specifies whether each such element (i) contributes primarily to the implementation of the Plan or (ii) contributes primarily to the achievement of other objectives but aids Plan implementation in important ways; and

"(B) states the portion of its request for appropriations that is allocated to each such element.

"(2) The Office of Management and Budget shall review each such report in light of the goals, priorities, and agency and departmental responsibilities set forth in the Plan, and shall include, in the President's annual budget estimate, a statement of the portion of each agency or department's annual budget estimate that is allocated to each element of such agency or department's high-performance computing activities. The Office of Management and Budget shall ensure that a copy of the President's annual budget estimate is transmitted to the Chairman of the Council at the same time as such budget estimate is submitted to Congress.

"ANNUAL REPORT

"SEC. 603. The Chairman of the Council shall prepare and submit to the President and Congress, not later than March 1 of each year, an annual report on the activities conducted pursuant to this title during the preceding fiscal year, including—

"(1) a summary of the achievements of Federal high-performance computing research and development efforts during that preceding fiscal year;

"(2) an analysis of the progress made toward achieving the goals and objectives of the Plan;

"(3) a copy or summary of the Plan and any changes made in such Plan;

"(4) a summary of agency budgets for high-performance computing activities for that preceding fiscal year; and

"(5) any recommendations regarding additional action or legislation which may be required to assist in achieving the purposes of this title.".

TITLE II—NATIONAL RESEARCH AND EDUCATION NETWORK

SEC. 201. (a) The National Science Foundation shall, in cooperation with the Department of Defense, the Department of Energy, the Department of Commerce, the National Aeronautics and Space Administration, and other appropriate agencies, provide for the establishment of a national three-gigabit-per-second research and education computer network by 1996, to be known as the National Research and Education Network, which shall—

(1) link government, industry, and the higher education community;

(2) be developed in close cooperation with the computer and telecommunications industry;

(3) be designed and developed with the advice of potential users in government, industry, and the higher education community;

(4) have accounting mechanisms which allow users or groups of users to be charged for their usage of the network, where appropriate; and

(5) be phased out when commercial networks can meet the networking needs of American researchers.

SEC. 202. In addition to other agency activities associated with the establishment of the National Research and Education Network, the following actions shall be taken:

(1) The Federal Coordinating Council for Science, Engineering, and Technology shall—

(A) establish a National Network Advisory Committee to provide technical and policy advice from all the interests involved in the Network program, including (i) researchers from university, industry, and Federal laboratories who will use the Network; (ii) university and college educators; (iii) librarians involved in electronic data storage and retrieval; (iv) industrial organizations that develop and provide relevant technology and services; (v) managers of regional computer networks; and (vi) experts in networking and computer science who can provide technical guidance;

(B) submit to Congress, within one year after the date of enactment of this Act, a report describing and evaluating effective mechanisms for providing operating funds for the long-term maintenance and use of the Network, including user fees, industry support, and continued Federal investment; and

(C) allow recipients of Federal research grants to use grant moneys to pay for computer networking and other telecommunications expenses.

(2) The Department of Defense, through the Defense Advanced Research Projects Agency, shall be responsible for research and development of advanced fiber optics technology, switches, and protocols needed

to develop a gigabit computer network essential for the Network.

(3) The National Institute of Standards and Technology shall develop, in cooperation with the National Security Agency and other relevent agencies, a common set of standards to provide interoperability, common user interfaces to systems, and enhanced security for the Network.

(4) The National Telecommunications and Information Administration shall determine to what extent current Federal telecommunications laws and regulations hinder or facilitate private industry participation in the data transmission field. Within one year after the date of enactment of this Act, the Administration shall report such determination to the Congress.

SEC. 203. In addition to such sums as may be authorized to be appropriated to the National Science Foundation by other law, there are authorized to be appropriated to the National Science Foundation for the research, development, and implementation of the National Research and Education Network, in accordance with the purposes of this title, $50,000,000 for fiscal year 1990, $50,000,000 for fiscal year 1991, $100,000,000 for fiscal year 1992, $100,000,000 for fiscal year 1993, and $100,000,000 for fiscal year 1994.

TITLE III—NATIONAL INFORMATION INFRASTRUCTURE

SEC. 301. The National Science Foundation shall coordinate, in close cooperation with the Department of Commerce (in particular the National Oceanic and Atmospheric Administration, the National Institute of Standards and Technology, and the Bureau of the Census), the Department of Defense, the National Aeronautics and Space Administration, and other relevant agencies, the development of a national science and technology information infrastructure of data bases and knowledge banks accessible through the National Research and Education Network referred to in title II of this Act. The infrastructure shall include, but not be limited to—

(1) a directory of network users;

(2) provision for access to unclassified Federal scientific data bases, including weather data, census data, economic data, and remote sensing satellite data;

(3) rapid prototyping of computer chips and other devices using centralized facilities connected to the network;

(4) data bases and knowledge banks for use by artificial intelligence programs; and

(5) provision for international collaboration among researchers.

TITLE IV—SOFTWARE

SEC. 401. (a) The Office of Science and Technology Policy, as indicated in the National High-Performance Computer Technology Plan (hereinafter referred to as the "Plan") developed and implemented under title VI of the National Science and Technology Policy, Organization, and Priorities Act of 1976, as added by section 101 of this Act, shall oversee the cooperative efforts of Federal departments and agencies in the research and development of high-performance computer software, including projects focused on astrophysics, engineering, materials, biochemistry, plasma physics, and weather and climate forecasting.

(b) The National Science Foundation shall establish clearinghouses to validate and distribute unclassified software developed by federally funded researchers and other software in the public domain, including federally funded educational and training software. Such clearinghouses shall--

(1) maintain libraries of programs;

(2) provide funding to researchers to improve and maintain software they have developed;

(3) help researchers locate the software they need;

(4) make software available through the National Research and Education Network; and

(5) promote commercialization of software where possible.

(c)(1) The National Science Foundation shall place special emphasis on the development of artificial intelligence and shall establish joint research programs among government, industry, and the higher education community to develop artificial intelligence applications.

(2) for purposes of this section, the term "artificial intelligence" means software and hardware which can be used for computer systems that learn, exhibit knowledge of themselves and their environment, make logical inferences, display creativity, or mimic other aspects of human intelligence, and such term includes expert systems, neural networks, natural language processing programs, translation programs, and higher level programming languages.

(d) The National Institute of Standards and Technology shall develop standards for software programs purchased or developed by the Federal Government that promote development of interoperable software systems that can be used on different computer systems with different operating systems.

(e) Procurement regulations at the Defense Department and other departments or agencies shall be changed so that contractors providing software to the Federal Government no longer are required to forfeit the proprietary software development tools that they used to develop the software.

SEC. 402. There are authorized to be appropriated to the Office of Science and Technology Policy for distribution to the National Science Foundation, the Department of Defense, the Department of Energy, the National Aeronautics and Space Administration, and other relevant agencies for computer software research and development, in accordance with the purposes of this title, $50,000,000 for fiscal year 1990, $100,000,000 for fiscal year 1991, $150,000,000 for fiscal year 1992, $200,000,000 for fiscal year 1993, and $250,000,000 for fiscal year 1994.

TITLE V—COMPUTER SYSTEMS

SEC. 501. The National Science Foundation shall ensure that the national supercomputer centers in the United States continue to have the most advanced, commercially available supercomputers produced by United States manufacturers.

SEC. 502. Where appropriate, Federal agencies shall procure prototype or early production models of new high-performance computer systems and subsystems to stimulate hardware and software development in the American high-performance computer industry. Particular emphasis shall be given to promoting development of advanced display technology, alternative computer architectures, advanced peripheral storage devices, and very-high-speed communication links.

SEC. 503. Within sixty days following the date of enactment of this Act, the Secretary of Commerce shall review export controls that hinder the development of foreign markets for United States manufacturers of supercomputers and other high-performance computer technology, and report to the Congress the results of such review.

SEC. 504. There are authorized to be appropriated to the Office of Science and Technology Policy, for distribution to appropriate agencies and departments as specified in the Plan, for research in computational science and engineering, $30,000,000 for fiscal year 1990, $60,000,000 for fiscal year 1991, $90,000,000 for fiscal year 1992, $120,000,000 for fiscal year 1993, and $150,000,000 for fiscal year 1994.

TITLE VI—BASIC RESEARCH AND EDUCATION

SEC. 601. The Office of Science and Technology Policy shall, in cooperation with relevant departments and agencies—

(1) support basic research on computer technology, including research on advanced semiconductor chip designs, new materials for chips, improved chip fabrication techniques, photonics, and superconducting computers;

(2) create technology transfer mechanisms to ensure that the results of basic research are readily available to United States industry;

(3) promote basic research in computer science, computational science, electrical engineering, and material science; and

(4) educate and train more researchers in computer science and computational science.

SEC. 602. To expand its traditional role in supporting basic research in universities and colleges, and in training scientists and engineers in computer science, computational science, and electrical engineering, there are authorized to be appropriated to the National Science Foundation, $10,000,000 for fiscal year 1990, $20,000,000 for fiscal year 1991, $30,000,000 for fiscal year 1992, $40,000,000 for fiscal year 1993, and $50,000,000 for fiscal year 1994.

○

Senator GORE. Very good.

Let us call the first panel up here now. John Rollwagen, Chairman and Chief Executive Officer of Cray Research, Incorporated, based in Minneapolis and Dr. Sheryl Handler, who is President of Thinking Machines Corporation in Cambridge, Massachusetts.

I might say to start what is already well-known and that is that Cray is the leading supercomputer company in the country and in the world and Thinking Machines Corporation is in my view the leading massively parallel computer company in the world. A different approach, and each machine does different things extremely well.

Without objection, your prepared statements will also be included in the record.

In introducing you, I want to say a word of thanks. Both of you cancelled your schedules in May to come and help with another similar event and both of you have juggled your schedules to be here today. I know you are both extremely busy CEOs and we are very happy that you are here today.

Mr. Rollwagen, we will start with you.

STATEMENT OF JOHN A. ROLLWAGEN, CHAIRMAN AND CHIEF EXECUTIVE OFFICER, CRAY RESEARCH, INC.

Mr. ROLLWAGEN. Thank you, Senator, and I want to express my gratitude, really, for having an opportunity to speak to your Subcommittee about this issue. I want to compliment both you and your staff and the Subcommittee generally for taking up this topic.

Obviously, I have a special interest, but I thoroughly believe this is important for the country as a whole and a very significant issue and very appropriate for you to deal with.

Senator GORE. Could I ask you to move that? It is kind of a funny mike. If you could move it closer.

Mr. ROLLWAGEN. In proceeding, as you indicated I have prepared a statement and I will not bore you with reading that, but I would like to make a few comments.

Before I do I would like to indicate as well that I am pleased to be able to speak both as CEO of Cray Research and as a representative of the American Electronics Association. We have had some discussions within the AEA itself and my statement can be interpreted as coming from that association as well.

What I would like to do in taking advantage of my time is give you just a brief background of the supercomputer business from our perspective, bringing it up to the historical point we are today.

I also brought a videotape which I think might help illustrate some of the points that you were talking about, Senator, as regards to the power of the systems and what they mean for scientists and engineers today and particularly in the future.

Then I will say just a couple of things about why we support the particular bill that has been introduced from Cray Research's standpoint.

The first point I want to make is that the supercomputer industry is a home-grown industry. It started in the United States and it is still dominated by United States suppliers. However, the character of the industry has changed considerably in the time since it first originated.

It is hard to put a specific pinpoint on that date, but at least in our terms our company started in the early 1970s, developed and delivered its first system in the mid-1970s. At that time we could identify precisely 86 potential customers in the whole world for this technology.

While that was enough to start a company, it certainly was not enough to generate the kind of interest that supercomputers have today and we would not be having any Committee hearings in the Senate of the United States to talk about a market confined to 86 potential customers.

Something has happened in the meantime that has changed the character of the business. First of all, just in terms of the current activity in the industry, rather than 86 potential customers, just our company has about 170 customers of those original 86, and we are talking probably to another four or five hundred companies ourselves, not to count the activity that Sheryl is involved in and the other companies that are in this business.

So we are at least an order of magnitude beyond where we thought the business would be in the early days and that is very exciting and very gratifying for us.

But even that is not enough to justify or warrant the kind of attention that supercomputers are generating, because I have the feeling that today we are as much on the threshold of this business as we were in the middle 1970s. It was exciting then. It is even more exciting now.

The reason I feel that way is that the technology has led us into a new world from a couple of different perspectives. First of all, the computers themselves have gotten powerful enough that the simulations they perform, which is what they are used for, is to simulate physical phenomena, are much more realistic and much more accessible than they used to be in the past.

We can do simulations now in three dimensions instead of two dimensions and as a result the simulations themselves are more informative, more realistic, more useful.

Secondly, there have been other devices and capabilities, complimentary capabilities, developed particularly in the area of work stations, with graphics capability that allow researchers to deal with these machines in the form of pictures rather than numbers and equations exclusively.

This means now for the first time that scientists and engineers can use this technology in terms that are intuitively comfortable. It is not necessary to be a sophisticated mathematician or a sophisticated computer scientist to get the advantage of computer simulation in doing engineering and scientific research.

Now, all of this still may sound rather, almost mundane or interesting but kind of techy. To put it into the final perspective I want to quote something that Paul Gray, who is the President of MIT, told me. He said it and I have heard it elsewhere as well.

In talking to me a year or two ago, he said, John, you may not understand what is happening as a result of the technology that you people and others like you are delivering.

He said, this is a very fundamental change in how science is pursued. In fact, nothing like this has happened in 300 years. The scientific method as we know it today really was born in the 1600s with the mathematics of Descartes and the empirical observations of Newton, of course, and the proverbial apple that landed on his head. Those are the two basic paths to scientific discovery that have been used since that time and the numerical simulation represents a whole, new third path towards scientific discovery.

Now, if that is true, that is a very, very fundamental change in how engineering and science is done and it certainly warrants the interest of the Senate of the United States and in fact any country, any organization that has an interest in participating in the development of science world-wide. So again, I think you are on the right track and it is something that is very important to pursue.

I think I can give you a little more substance to what I am saying if I show you that tape right now, and I wonder if I could ask you to tip that screen just a little bit this way because I have to kind of watch what is happening and this does not have a narration. We are kind of primitive in our PR but at least you will get the idea what these pictures are.

The first one is a weather application that was developed at the National Center for Atmospheric Research. It will show you, as you see, a map of the world laid out in front of you. In the red are high pressure areas. Blue is low pressure and green shows the jet stream, and this is at a point three and a half miles above the earth.

Now if you think about it, that is a pretty monumental thing to be able to picture and this is all the product of equations and numerical analysis that is being done on the computer and that allows the atmospheric scientists to study, for example, the greenhouse effect and a lot of the work on that was done at NCAR.

This is a fascinating one done by Dr. Gregory McKay at Carnegie-Mellon and this shows the Los Angeles Basin with the base case on top, which is what really happens every day in Los Angeles.

You are going to see some red begin to appear in the basin and that represents primarily ozone, but other chemical pollutants engulfing the area and we all know that happens in Los Angeles.

He was able to run simulations successfully on one of our systems that showed if we could convert from gasoline to methanol in engines, automobile and truck engines and so forth in the Los Angeles Basin, that in effect the problem could literally be solved.

This is having a major impact on policy decisions in that area and it took literally hundreds of hours of computer time to come up with that simulation. I know that President Bush when he was Vice President was made aware of this and got very excited about it in his responsibilities for the EPA.

The next one is even more fascinating, I think. It shows a tornado but we will have to wait a minute until we get there. Now this looks like a cartoon but it is not. This is the output of a program to simulate the activity of a tornado, presumably in Kansas or some place as you can see by the field. Pretty soon that cloud will come to life, as you can see, and start moving.

As I say, it looks kind of like a cartoon, but no artist drew this picture. This is the output from the computer as it simulates—and you can see the tornado forming right there—as it simulates how tornadoes actually work and progress and move over the landscape.

Now, one of the fascinating things about this is that it illustrates the point I was saying about communicating with pictures instead of numbers and equations. If the computer printout for this program were just printed, it would be three-and-a-half feet thick, and try to make sense out of a three-and-a-half foot computer printout, rather than that picture.

This, as you can probably tell, is an F-16—oh, it is written on there—and it shows how the air flow works over an F-16 and it is a direct illustration of what you were talking about, Senator, with the use of numerical simulation instead of wind tunnels, and in fact it illustrates how this is even better than the wind tunnel because the wind tunnels do not draw those pretty lines.

It is very hard to interpret the data from the wind tunnel. This is similar, only this is the National Aerospace plane, or one version of it, and in this case what you are going to see it is going to be flying at mach 25 and white is high stress and black is low stress. I guess the black is not showing up very well. But that has great significance for the designers of that plane and to be able to deal with that output.

The next one is another aerospace application. There is a real problem if you want a jet plane to take off vertically in terms of, where does the exhaust go after it hits the ground and that has to be engineered very carefully so as not to come back into the intake into the engine, which would be disastrous for the plane and the pilot.

These look very straightforward, these lines and so forth, but the underlying physics and the equations to run that are very complex and require the power of a supercomputer.

The next two are particularly fun. They are from the automotive industry. The first one will show you a connecting rod. The two designs are the same here, but in this case the computer was given the assignment of reducing the weight of the connecting rod and maintaining its physical integrity.

As you see, it is changing shape and it is illustrating where the stresses are again with the colors, but take my word for it—or take the Cray's word for it—the one on top has all the strength of the one on the bottom, and based again on the simulation of the physics involved, and it would make an obvious difference in the cost of that particular part.

Even more impressive than that is the next simulation which we have actually talked about quite a bit, which is a car crash. Now, this is the thing and I want you to watch this. It is a real Volkswagen Polo being crashed into a real wall and just watch what happens. We have all seen these kinds of films. Particularly watch it bounce back right there, which evidently is a very significant thing in automobile engineering.

Here is the computer version of the same thing and this was used to decide whether to get one of our computers, and the thing that they were looking for is, did it bounce back? If it did not bounce back, then the simulation was not real. Thank God, it bounced back.

We did not know that in advance. They did not just program that in. But you can see how it is very similar. For practical purposes, identical to the real thing and yet running this on a computer means that you can change the design instantly and experiment with it. The only problem is, that simulation took 11 hours to run on a Cray, so even the Cray computers are not fast enough.

I will show you this last one just for fun, but it also makes a point. I do not know if you saw the movie 2010, but this is a picture of Jupiter that was used in the movie 2010, and with the spot and everything.

What is interesting about this is that that is a huge picture on a movie screen with great resolution because the Cray was able to draw the picture. But it was not an artist's rendition which was then animated on the computer. That part was—there is no hole in Jupiter like that. We disavowed that part of it. But that is a little bit of Hollywood. You can turn it off now, if you will.

The point I want to make is, if you saw that picture on a screen you saw perfectly high resolution, or a perfect picture of Jupiter, but the picture was created by simulation of the atmosphere on Jupiter.

We have those equations at the jet propulsion laboratories and the producer of the movie got them and ran them on the Cray and so that is a weather forecast for Jupiter, and there is every reason to believe that that is an accurate weather forecast and therefore an accurate representation of Jupiter

Because the picture is high resolution, that may be the most real picture of Jupiter that we have ever produced, including satellite fly-bys, because the resolution of those pictures are not high. So this may be a more accurate—a more real, if I can put it that way—picture of Jupiter than a real picture of Jupiter, which is a little strange to think about.

That is basically what the business that we are in now, is providing that kind of simulation and making it available, literally. If it is true, if this is a third route to science, if these are real simulations, if this is a meaningful way to do science and engineering, it is my contention that in the 1990s every single engineer and scientist in the developed world will use a supercomputer of one kind of another almost everyday in his or her work.

So we are not talking about 86 potential customers, or even 860 potential customers. We are literally talking about millions, perhaps, of potential customers using this in their daily work.

Now, a lot of things have to happen for us to realize that potential. Certainly the computers have to be available and the software has to be available to run the computers, but we have to be able to deliver that capability to the scientists and engineers in a form that is useful and easy for them to use.

That is where I think this bill comes in, in many different ways. As the Senator said, we are talking about the title that has to do with the network that is key for delivering this capability across the country to the hands of scientists and engineers all over this country and to do it effectively we need wide band width. We need a high capacity piping system.

But also, that is not even good enough. What we need is a collaborative effort among universities, industries, government laboratories, to share the knowledge that we are developing on how to create these simulations and how to use them and that is a key part of this bill as well.

We need to emphasize the education—scientific education—of our new scientists and engineers to know how to use these, both to know how to program them for the computer scientists, perhaps, and to know how to use them if you are just a regular engineer or scientist, effectively, and that is dealt with in this bill.

I think it is a very exciting opportunity we have to take an important step forward and to recognize—this is my final statement—to recognize that it is critical for us, because we are dealing in a world-wide competitive struggle to develop U.S. technology and to maintain ourselves at the forefront, not just of the supercomputer business and what that means, but for all of industry, really, that has anything to do with physical phenomena, any kinds of products and services that are built on science and engineering technology of any kind. This is a critical tool to make available to those people.

So—I could go on for hours. I will not. I commend you on your interest and would be happy to answer any questions later that you might have.

[The statement follows:]

STATEMENT OF JOHN A. ROLLWAGEN
CHAIRMAN AND CHIEF EXECUTIVE OFFICER
CRAY RESEARCH, INC.

also on behalf of

THE AMERICAN ELECTRONICS ASSOCIATION

Good morning, Mr. Chairman. I thank you for the opportunity to testify before your subcommittee on the networking portion of your "High Performance Computer Technology Act of 1989." Let me just say that Cray Research, Inc. appreciates the leadership you are giving to this very crucial industry and its technology base. My testimony today will also reflect the views on High-Performance Computing of the American Electronics Association (AEA) of which Cray Research is an active member. The AEA is the largest national trade group representing electronics and information technology companies. AEA membership includes 3500 companies and 48 major universities as associate members.

The national focus and discussions on your Bill will contribute to a broader understanding of the overall benefits to be derived from computational science. Specifically, it will accelerate the use of supercomputers in the pursuit of higher quality science and technology and will motivate the behavioral shift on a national basis to a computational approach to solving science and technology problems. We believe that this shift will put the United States on a new technology curve altogether and return greater benefits.

We at Cray Research, Inc. are very supportive of this legislation. It addresses several trends that we see as important, if not crucial, for the U.S. to increase the distance between our supercomputer industry and that of Japan. First, all sectors of the U.S. infrastructure must work together in a more collaborative fashion. Traditionally, programmatic efforts dealt with one sector or another; rarely did legislation deal with this critical cultural change to intersect collaborative efforts. While the U.S. has the largest R&D expenditure in the world -- at last count $132 Billion-- we work hard to separate our policies and funding activities based upon basic or applied research, advanced development, etc. Even further, we have separated our policies and our support differently for the university, private and defense sectors. This has resulted in a distortion of resources allocated for R&D important to the commercial sector and to overall competitiveness in world markets. In the past when global competitors were more evenly matched in terms of size and resources, and when company resources were the only resources in the private sector, these distinctions apparently had little impact. We, in the U.S., have had a phobia of an emotionally laden term known as "industrial planning" which has been translated to mean anything relating to government support for non-DoD supported industry. This has been true even when there has been clear evidence that structural differences between our infrastructure and other trading blocs would have long-term deleterious effects for U.S. economic security and so, too, national security. This legislation addresses this crucial need to break down these artificial barriers in order to be able to "communicate" across sectors. A "clear channel" is necessary to speed the transformation of important results, paid for by our R&D dollars from any one of the sectors, to industrial relevance. Japan has tremendous strength and well-established effectiveness in this very basic requirement of rapid, clear and effective technology transfer.

Second, the industries and the macroeconomic policies that support the supercomputer industry are significantly different between the U.S. and Japanese countries. There is no value judgment here; the structures are just different. In the U.S., we have our anti-trust mentality and we espouse "small is beautiful." In Japan, it appears that the notion of critical mass required to effect market-share oriented trade policies are de rigueur. That is, there are significantly less companies competing for the same market dollar and therefore, there is more market dollar available to fund enhanced R&D. In order to leverage the larger U.S. investments in R&D necessary to minimize the negative impacts due to mere structural differences, the technology base developed within any of the sectors must be able to complement or leverage the technology investments of the others.

This legislation appears to directly address this critical issue. Industrial participation is explicitly described. The model and experience our country derives from the implementation of this legislation will have benefit beyond the supercomputing industry. Using the technology investments in a manner that is indifferent to which sector developed the technology will begin to balance this structural disproportality between the supercomputing industry of the U.S. and that of Japan. Just by way of reference, Hitachi's revenue for 1988 was $23.9 Billion, NEC's was $18.9 Billion and Fujitsu's was $15 Billion, based upon a conversion rate of 130 yen/dollar.

Today, the U.S. has well-established leadership in advanced technology in the following areas: successful implementation of new architectures that exploit massive numbers of processors in parallel; significant software technology necessary to exploit high levels of parallelism; a deep bed of experience in algorithms necessary to address the leading-edge problems of science and engineering crucial to the economic and national security health of the nation; and a well-proven and successful vectorization capability. The Japanese have yet to deliver a multi-processor supercomputer. They have identified software as a crucial capability to acquire. They send significant numbers of researchers to study at our universities as well as to NASA and NIH to learn the computational methods necessary to fuel high-tech applications that can affect the competitive posture of whole industries. Continued and more aggressive investments in these technology areas by the U.S. will be significant to enabling "break-away" achievements for the supercomputing industry. The new thrust described in the legislation of including industry and focusing on industrial needs and issues will help ensure that the U.S. supercomputing industry can continue its leadership in the global marketplace. This legislation focuses on software as well as networking. These two technology areas of the Bill, in our opinion, will have the most direct impact on the supercomputer industry and on the high-tech industries that can leverage supercomputing into a globally significant competitive edge.

The supercomputing industry will benefit from this legislation because it will provide researchers better access to supercomputers and supercomputing technology, such as that available from the NSF Supercomputing Centers. A decade ago, few people would have predicted the impact that supercomputers would have on our lives. Today researchers around the world are applying supercomputer power to the solution of diverse and complex problems. With just one supercomputer system, users can solve a wide variety of problems. For example, in the aerospace field, supercomputers are being used to design engines, to analyze structural responses to stress, to design electronics system and to simulate wind-tunnel testing. Other application areas include: weather and atmospheric modeling, seismic processing for oil exploration, pharmaceutical and chemical, image processing, finance electronics and nuclear power. Though the impact of supercomputers has been phenomenal in the last decade, it is clear that the changes have only just begun.

It should be noted that this initiative will help educate more students and industry researchers about computational science in our nation's colleges and universities. For some time, the United States research community has been concerned about the education and training of America's current and future scientific and technical work force. Enhanced science and engineering education definitely must be considered as part of any long term foundation that must be laid in order for the United States to maintain its technological world leadership. This Bill is a specific measure that will improve the training of our scientific and technical work force and expand the supply of highly-trained scientists, engineers and technicians by providing more supercomputer access to individuals in the academic areas. Also, industries that depend on leading edge R&D will benefit by having greater access to these state-of-the-art supercomputing tools in addition to having greater access to the researchers from universities and the national laboratories who are working on related research.

It is clear that the supercomputer highway will quickly catapult U.S.-based researchers into new levels of productivity and creativity by providing not only more accessibility but also more timely results. It will also ensure that the computational approach will move quickly to take its place along side of experimentation and theory in the pursuit of new heights of science and technology. This will provide for greater rates of innovation - the only true offensive weapon in our economic race for global industrial leadership and continued national security.

The National Research and Education Network (NREN) is the key to addressing this problem by providing greater access to supercomputers by various industries. Your analogy to an "information super highway" is very apt. We also believe that the ultimate benefits will be beyond what we can envision at this point in time, as we begin to shape the beginning of our effort. What is clear is that the network is a crucial enabling function. Furthermore, the NREN builds on investments already made by the United States Government, where benefits have clearly been achieved: ARPANET, MFENET, NASANET, SURANET, etc. It has been wonderful, indeed, to watch the networking activities at local levels, even on university campuses, as researchers seek out fellow researchers and distributed computational resources important to solving their research problems are brought to bear on specific areas of investigations. This has no doubt been a strong contributor to the continued excellence of our university system.

Our government has acted as a catalyst in releasing all this grassroots energy: NSF has already "built" the network backbone for NSFNET; NASA has built effective networks to U.S. industry; universities have built "routes" into supercomputing centers as well as into the national laboratories to allow researchers to access state-of-the-arts tools of R&D and to communicate with the resources available throughout the country. It now seems appropriate in this stair-step process that the government would establish an enabling framework to allow all these diverse and somewhat uncoordinated grass-root activities to work in a more effective coherent and productive manner. This enabling framework is the NREN. By adding significant capacity as envisioned by your Bill, some of the barriers to solving problems remotely will begin to be addressed. Today, I understand that some researchers choose between taking their data on an airplane to a supercomputer center or submitting the data at their local work station. If the research requires a large bandwith to handle the problem data, the airplane is many times the most realistic option. Many leading edge problems of science and engineering in fact require not only many computer cycles in order to obtain the quality of science envisioned, but also are characterized by huge quantities of data. Weather forecasting is an example where huge quantities of data need to be computed in a timely fashion. Forecasting the weather for the next several days, assessing the influence of the constant interplay between the atmosphere and the oceans on the coming winter weather, and predicting the climate consequences of increasing carbon dioxide in the atmosphere, all depend upon the lighting speed and large memories of today's supercomputers. Obviously, if one has to take data by plane to the computing center, much of the spontaneity that we intuitively associate with creativity and achieving higher levels of science is lost.

The proposed legislation moves the U.S. closer to our goal of becoming better at "commercialization of the results of our R&D" through better technology transfer and dissemination. The National Network Advisory Committee identified in the Bill, consisting of representatives from all sectors, should ensure more effective collaboration among researchers no matter what sector they represent. The transfer of technology among these sectors should also be accelerated. Having the Advisory Committee undertake deliberations of difficult issues will ensure that workable policies emerge, further ensuring the success of the Act. Although we recognize that the nation faces real budgetary constraints, we would like this Committee to consider having a window of time - say 2 years - during which researchers from all sectors would have "free" access to networks. This should be coordinated with access to low cost supercomputing resources in either the national laboratories or the NSF Centers. The U.S. is very much at the beginning to what amounts to a culture change --in the jargon today -- a "paradigm shift" to computational science. In order to accelerate this shift - necessary to reap the benefits anticipated by the Bill - all barriers between sectors should be removed in so far as is practical. Fees for access and supercomputer use represent very real barriers and will work against the goals of the Act.

From the perspective of a vendor that sells supercomputers throughout the world, we have witnessed an interesting scenario occurring in the way supercomputers are being used in other countries, when compared to the way they are used in the United States. Japan, even perhaps more than the U.S., sees supercomputers as an essential tool of R&D. Japan has a larger industrial and university commitment to supercomputing than we do and they apply their supercomputing technology primarily to enhance the competitiveness of their commercial industries. Of course, we believe that supercomputers are important to the competitiveness of a broad range of U.S. industries. However, Japan has significantly more supercomputers in application areas which could be termed "competitively embattled" than the U.S. Japan uses more supercomputers in automotive, chem/pharm/biotech, construction, and electronics. The U.S. uses more supercomputers in aerospace, petroleum, and research, traditional areas of the U.S. strength. Take the automotive industry for example. The U.S. big three car makers have all purchased Cray machines. However, most of these machines are our smaller and therefore less powerful supercomputers. On the other hand, our European and Japanese automotive customers have stepped up their purchases of Cray's larger multiple Central Processing Units (CPU) and therefore more powerful systems. These computers can be used to design almost every component in a car as well as be used for simulated car crashes and thereby avoid building multiple prototypes. Of the 15 Crays used by the automotive industry, three are in Detroit, compared with seven in Europe and five in Japan. Honda, Toyota and Mitsubishi have one each. Nissan's two are so powerful that they outperform our nation's big three combined. The message here is simple - our competitors around the world recognize the advantage they can receive by greater access to supercomputers and make this a priority in their R&D efforts. In a recently completed study performed under the Japanese technology assessment program (JTECH), on computer integrated manufacturing for Japan's semiconductor industry, it was noted that Japanese companies routinely use supercomputers and that supercomputers are seen as necessary tools of R&D. There are 9 supercomputers in use in Japan in the electronics industry compared to 3 in use in the U.S.

I might also add a thought at this point about the NSF Supercomputing Centers. While we are very excited about the NREN, it seems that additional specific attention should be given the NSF Supercomputing Centers. It would seem that increasing the capacity of the networks into the Centers will also significantly increase the requirements of the Centers to support their increasing base of users. The NSF Supercomputer Centers have played a unique role in the cultural shift to computational methods and support for them should be expanded to ensure their continued role as leading edge technology centers. The Centers provide state-of-the-art supercomputing facilities; they are a research repository for our largest problems -- setting the pace for other Centers established by the states and elsewhere. These Centers have been successful in disseminating supercomputing technology into a broad user community. Also, new "programs" that will allow these Centers to continue to "set the pace" in such areas as visualization, broader training, novel national grand challenges, etc. would ensure that the goals of the Act are effectively and aggressively realized. At any rate, it would be our hope that the implementation and funding of the Bill would provide recognition that the upgrade schedule for the Centers could change due to an anticipated increase in the number of new researchers and that special activities, such as training and support, might be necessary and should be put in place proactively.

In addition to the network, Cray Research believes that the other high payoff area of the Bill is software. This includes software tools for debugging compilers that are better at detecting parallelism from code; parallel algorithms; and the development of new application codes for leading edge science and engineering. We believe that the benefits from an intense focus in the software area will have high payback to a broad range of researchers. We would be glad to explore this and other areas of the Bill in more detail at another time.

Thank you again for this opportunity to testify.

Senator GORE. Thank you very much. We will hold questions until Dr. Handler has spoken.

Dr. Sheryl Handler is the CEO of Thinking Machines Corporation in Cambridge, Massachusetts. Thank you for being here.

Please proceed.

STATEMENT OF SHERYL L. HANDLER, PRESIDENT, THINKING MACHINES CORP.

Dr. HANDLER. Thank you.

Today I would like to talk about the supercomputer industry from a slightly different point of view, both because our technology is different, but also I think our perspective is different. I would like to talk about why the country that leads the world in high performance computing will also lead the world as the high performance economy.

We are in an age of economic, not military, confrontation, and high performance computing is the essential ingredient in leadership in both business and science. ~~And tha~~t is my key point today, that the supercomputer industry is important in its own right, but it is an enabling technology that will have a very significant impact in business and in science in a much larger way than in the past.

Leadership in the supercomputing industry takes two forms. First, it means leadership in the manufacturing and sales of supercomputers worldwide, but it also means leadership in supercomputer usage and software in the United States. And the proposed National Network will provide an important boost to software development and utilization in America.

Today I would like to talk first about the importance of supercomputing to the economy. Then I would like to talk about our current lead in a new wave of supercomputing. And then I would like to describe how we can capitalize on this lead in this new technology.

Also, I have a cold, so let me know if you cannot hear me. I will try to speak louder.

First, the importance of supercomputing. There has been a lot of publicity lately about the strategic importance of supercomputers, and it is all true. Those with access to the biggest supercomputers achieve the deepest understanding.

Two hundred years ago Napoleon said that God is on the side of the largest army. Today, he would just as accurately say that God is on the side of the largest supercomputers. They are that important.

Supercomputers solve the biggest and the hardest problems. First, they help us to make use of our largest data bases. We have spent years accumulating data and now find that it is very hard to have access to it. With today's new supercomputers, we find this wealth of data usable in a practical form for the first time.

These new supercomputers find, in effect, the needle in the haystack. They can access a single document from amongst millions. For example, physicians could search the entire National Institute of Health medical data base looking for rare disease diagnoses and treatment protocols, or businesses could search and pinpoint statistical and demographic data necessary to support new product decisions and changes in customer demand. Scientists could find the results of experiments done halfway around the world. The existence of a National Network will assure that the benefits of America's growing data bases are available to all.

Now, supercomputers can also do the reverse. They can see the forest through the trees. They can spot trends in millions of pieces of data, and this capability is really vital to businesses as they compete on an expanded worldwide basis.

Now, just to underscore the importance of data bases, I would like to remind you that when an aerospace company delivers a jumbo jet, the documentation it delivers weighs more than the airplane.

Senator GORE. Is that true?

Dr. HANDLER. That is what one of our customers has told us, and we believe our customers. That is a lot of documentation, I know.

The second way supercomputers contribute is by helping scientists to understand nature. It is hard to understand an ocean because it is too big. It is hard to understand a molecule because it is too small. It is hard to understand nuclear physics because it is too fast, and it is hard to understand the greenhouse effect because it is too slow.

Supercomputers break these barriers to understanding and, in effect, they shrink oceans, they zoom in on molecules, they slow down physics, and they fast-forward climates. And clearly, a scientist who can see natural phenomena at the right size and at the right speed learns more than one that is faced with a blur.

Now, how will the National Network help to make these benefits more available throughout our economy? It will do it by making supercomputer access independent of supercomputer ownership.

Anyone anywhere in America who needs access to data bases or needs to simulate an event could potentially do it over the proposed network. The network is particularly important now because the world of supercomputing is in the midst of a transition. Until recently, supercomputers have been very expensive and, therefore, relatively few have used them. Now that state of things is changing, and therein lies the biggest opportunity for America.

The new wave of computing that is emerging is many times more powerful and is also less expensive. What is this new wave? It is known as parallelism. The highest performance version of parallelism is called massively parallel supercomputing, and the concept is very simple. If you have a problem that has millions of pieces of data, the fastest way to compute it, the way nature does, is to compute all the data at once. This is true for both business and scientific applications.

America is ahead in this new form of computing. That is because America, through DARPA, took the risk five years ago to try new approaches, and it paid off. We have built and begun using these machines, and we are way ahead of the foreign competition.

Massively parallel computing enables us to achieve performances of more than a billion operations per second, a gigaflop, at prices starting in the hundreds of thousands of dollars. This makes supercomputing affordable by those who now buy minicomputers and mainframes. So for the first time in the history of computing, massive parallelism is enabling us to continue to build computers that are increasingly faster and simultaneously bring the price down.

In the past 50 years new markets have emerged by bringing the price down first—mainframes to work stations, work stations to PCs—but in each case, the power of the computer was reduced. But now we can build machines that are ever faster and also bring the price down.

More powerful supercomputers means that bigger and tougher problems will be solved. Cheaper machines mean that more users will be able to harness the power of supercomputers. So, in sum, more will be done and by more people. This is very American, but the question is, will America take the next step.

I would like to turn to the future. America is the unequivocal leader in parallel supercomputing. We know how to experiment and innovate, but are we going to capitalize on our investment by building a dominant market position both here and abroad. Or, like high definition TV, are we going to let our technology slip away to others who see the economic potential and make a commitment to it?

What is the reason this might happen? The reason is that we are faced with a double transition. One transition is that we must now adjust to a highly competitive environment. America is used to having a virtual monopoly on supercomputing, but that has ended. And the other transition is that we must move from traditional supercomputing exclusively to having traditional supercomputers and parallel processing.

The good news is that while we are battling it out with foreign competition in one area of the industry, we are leaping ahead in another. The bad news is that Europe and Japan are not far behind.

Many countries recognized the potential of parallel computers very early. Even before the United States, Japan and Europe had large government programs to build these new machines. American innovation prevailed nevertheless, and we got to the goal first. But the foreign appetite is very strong and sophisticated, and they are determined to be major players in this field.

We, on the other hand, are so used to being the leader in supercomputing that we do not have the sense of urgency about using new parallel supercomputers that other countries do. As an example, many of our leading supercomputer centers have not committed to new architectures because they are waiting for the software to mature. By the time this happens, we will find ourselves buying the hardware overseas from those who set out to win.

The Federal Government must make it clear that we cannot, cling to traditional supercomputer technologies while others adopt the newer wave of parallel processing in order to leapfrog us. We should not abandon our traditional supercomputers, but we must embrace this new wave while it is emerging. The proposed National Network will help America to adopt this new technology faster by making it available over the net.

In order for America to solidify its lead in parallel supercomputing, we must do two things better than we have in the past: We must get American users onto the new technology very quickly and broadly. We must export overseas with great gusto.

We make the following recommendations:

First, the Federal Government should move quickly to install the 3-gigabit National Research and Education Network. It is a vital component of the infra-structure of an Information Society. The information age will not be here until you can access the Library of Congress from your desk.

Second, as a matter of policy, the Federal Government should assure that the new parallel computers are installed on the Network as soon as they are available. The Nework should serve to foster American innovation.

Third, the Federal Government should provide priority access on Network to users who commit to write new software. As new software is written, it should be made available to other users over the Network.

I think, in summary, I would like to point out that supercomputers, once again, are used to solve the biggest problems and the hardest problems. Our country cannot afford to lose its competitive edge in high performance computing. Either we learn to take quicker and better advantage of our new supercomputers, or they will not be built in the United States any longer. It is that simple.

This country must be the first to have every scientist, every businessperson, every student using supercomputers. The National Network is a major step toward this goal. If we do this, we will do better science, we will operate our businesses with more information, and we will make better products. And this is what it will take to be the high performance economy of the 1990s and beyond.

Thank you.

Senator GORE. Thank you very much. Thank you both for excellent statements.

Before I start the questions, I want to note the arrival of our ranking Republican.

Senator Pressler, did you have any comments?

Senator PRESSLER. I would just say that in establishing a national high speed computer network, we have to make sure that the small schools and schools in the rural states are not left out, obviously. Today, access to computer networks is limited by location or funding. Faculty from smaller schools and rural states such as I represent—and in part you do, Mr. Chairman—need to be able to use our Nation's unique research resources such as supercomputers and weather data bases and so forth. This will enhance the schools' research and education programs and contributes to the Nation's knowledge base.

A computer network across the United States obviously has great potential, and I am happy to join in supporting this effort. And I hope that the Bush Administration makes a clear statement of support, which they have to some extent.

Senator GORE. I am going to suggest that we follow the five-minute rule in this hearing, because we are blessed with very good attendance, which I think is great. But I am going to start by trying to keep within five minutes myself here.

On the question of the access to rural areas and smaller universities, I do share that concern. The limiting factor today is—and we will hear from other witnesses later on this point, and they can correct my impression if it is wrong—the real limiting factor is the availability of the electronics, the switches, the software that enable us to use the fiber optic cables that are already in place.

The opening up of the long distance telephone market to competition meant that MCI and Sprint and a dozen smaller, usually regional companies laid lots of cable with huge redundancies in their inherent capacity. The development of the switches, the software, the electronics for the National Network will make it possible for that same technology to quickly be used on existing fiber optic cables in South Dakota, and in other parts of Tennessee, and in Mississippi, and in Virginia, and in Nevada.

Have I covered everybody here?

In any event, I think that that is a problem that will be taken care of.

Mr. Rollwagen, can you give some specific examples in dollar terms of how supercomputers have made a business more profitable, more productive, or more competitive? Do you have a specific example that we can think about?

Mr. ROLLWAGEN. Well, there are several. One that comes to mind when I am asked that question always is an early one. One of the first users in the petroleum industry was ARCO that got one of our systems. It may even have been in the 1970s or the early 1980s. And within six months of having the system, they figured out by simulating that oil reservoir in Prudhoe Bay, they figured out how to get 2 percent more oil out of the ground than they could have if they had not been able to simulate the reservoir and the various methods of extracting oil. And 2 percent does not sound like much, but that is on a base of at least $100 billion, so there is a $2 billion benefit that accrued to ARCO at the time. And I remember trying to negotiate 2 percent of their 2 percent, and they said forget it; we have already bought the machine for $12 million.

There is also a case—and the numbers are not coming to mind; maybe somebody here from the Cray staff has got it—but there is a case within the Materiel Air Command in the Air Force of being able to forecast the weather more accurately, which they do now with one of our systems. And they save literally millions of dollars a year in fuel costs, simply in fuel cost of moving material around for the Air Force.

I know that Grumman, when they designed the latest version of their Gulfstream plane, came up with a new wing in a record amount of time. I am not sure what the total economic benefit was. I know that Boeing—maybe you have noticed on the initial, or on the new 737 of Boeing, their new version of it, it has an engine with an odd shape in the front. It has a triangular opening on the engine rather than a round one. And the placement of that engine and the design of the engine was revolutionary enough that they could not have justified that without having simulated it first on a Cray. They could not have gone through all the process of the physical experiments, but they were satisfied with the simulation such that they went ahead and built the plane, and lo and behold, it produced something like a 30 percent fuel savings in that design. So there is story after story.

Senator GORE. Well, you can elaborate for the record. We would welcome you doing that.

Now, the reason we saw so many pictures in your demonstration has to do not so much with the way the computer naturally presents the information as with the human mind, correct? I mean, the reason you use pictures is because that is the way the human mind can best absorb lots and lots of data, is that right?

Mr. ROLLWAGEN. That is correct. One way to think of it—and it is a technical term, but I think it has some significance if you are familiar with it—there is a tremendous impedance mismatch between us and machines, between us and computers. We do not think in the same terms, so to speak, except maybe Sheryl; she is working on that.

In any case, between our computer and human beings there is quite an impedance mismatch. Then switching to pictures matches that impedance. Again, that is the form in which we think and which we can absorb massive amounts of data.

Remember that one, the thunderstorm example. We could all absorb and understand, comprehend what was going on there. I defy any of us to read three and a half pages of computer output and make any sense of it whatsoever.

Senator GORE. If you try to analyze the human mind in the terms used to describe computers, you might say the human mind has a low bit rate, if you try to absorb information bit by bit by bit, but very high resolution. If we see a mosaic or a pattern or a picture, we can instantly absorb lots of data just by seeing how the different bits of information relate to each other.

And supercomputers offer, for the first time, the ability to organize and present massive quantities of data, not just as rows and rows of numbers, but in a form that makes it possible for us to understand what is going on.

Mr. ROLLWAGEN. And then the interesting thing is, having absorbed all of that, it immediately stimulates our own creativity. Oh, I had not realized that. I did not think of it that way. Now I want to do it this way. And it creates a demand for 10 times or 100 times as much information in return.

Senator GORE. And that is also the reason why Paul Gray at MIT says it is a new basic branch of knowledge, in addition to inductive reasoning and deductive reasoning. Either you think of a theory and go out and test it in the real world, or you go out in the real world and you encounter facts, and you try to resolve them into a theory that explains them.

But supercomputing, for the first time, allows us to simulate alternative realities with such detail that we can test out what works and what does not and go about exploring the world around us in a very, very different way.

Dr. Handler, the leading machines now, both the massive parallel computers and the more traditional machines—although it is hard to use the word "traditional" with respect to Cray—they all cost a lot of money now. What would a top of the line machine cost today?

Dr. HANDLER. Our most expensive—our largest machine, which is also the most expensive—is in the order of $6 million or $7 million.

Senator GORE. And Mr. Rollwagen?

Mr. ROLLWAGEN. About $25 million for the top of the line.

Senator GORE. And they do different things, of course, in different ways. But in both cases is it fair to predict that within five years the cost will be coming down dramatically?

Mr. ROLLWAGEN. Let me, if I may, talk about what has happened in the last perhaps 10 or 12 years. The first machine we sold in 1976—excuse me, 1977—was sold for $8.8 million, called a Cray I. I have on my desk at home a new personal computer that happens to be from Apple that has a larger memory, and it has almost as much computational power, believe it or not, as that Cray I, and I am convinced that within the next couple of years I will be able to have such a machine for perhaps $8,000. So a factor of 1,000 to one in performance and in reduction in price, or 1,000 to one in reduction in price. At the same time we are delivering machines now that are measured in hundreds of times the performance of the Cray I, and we will clearly also develop a machine in the early 1990's at a thousand times the performance of the Cray I. And we see no end in sight.

Senate GORE. Dr. Handler.

Dr HANDLER. I think we see the direction going both ways—more expensive and less expensive at the same time. And it is not far off into the future that we will be able to build computers, supercomputers that are as big as utility plants and will cost hundreds of millions of dollars. And that same technology in smaller pieces will provide even more supercomputing power than we have today, right on your desk. So that they will be going both directions, and that is part of the breakthrough, that computing—we are going to have vastly more computing power, and we are going to need to learn how to use it; and if we do not, someone else will.

Senate GORE. I have not stayed within the five minutes. I just want to clarify the point I am trying to make with those questions.

Supercomputers are generally defined as the most powerful, best, fastest computers available at any given time, but machines that are capable of transmitting data usefully and productively at the rate of a billion bits per second, which now costs $10 million to $20 million, $7 million to $20 million, will very soon be in the hundreds of thousands of dollars range, and then even lower in price. Yet, we do not have the network capacity to allow American businesses, American researchers to talk with each other and communicate with each other as this development takes place. If we do create that capacity, then the synergy, the mutually reinforcing productivity that can take place all over this country, will be absolutely phenomenal, in my view.

But, again, I have gone overtime, and I apologize.

Senator Bryan.

Senator BRYAN. Thank you very much, Mr. Chairman. I think all of us in the Congress have been inundated this session with concerns about maintaining our competitiveness as a nation. There's been testimony before this subcommittee on the latest technology and on proposals for space technology.

A few years ago as governor many of us were importuned by the exciting prospect of the Superconducting Supercollider, and yet I think we all recognize that although we would like to have all of these things, that it unfortunately becomes the responsibility of the Congress to prioritize, to make some choices.

Let me ask you to be as objective as you can be. of the scientific options that we would all like to pursue, how do you rank supercomputing in terms of importance to our nation and to our future ability to compete in the global marketplace?

Dr. HANDLER. I think the important issue is that we don't have to make a choice in one sense, and we have to make a very careful choice in another.

The whole idea of supercomputing technology is it is an enabling technology that allows work to be done better in virtually every field, and that will become clearer over the next five years as more supercomputers are able to be used because they are more affordable.

The strength of America has always been in its diversity of approaches and for us to select one science that should get attention over another, I do not think anyone is smart enough to do that.

What America can do is provide the tools that let the aggressive and talented people in science and business exploit those tools as best they can. And I think it is very important to encourage and stimulate and provide incentives to get as much software developed for supercomputers.

If that's done, then in one fell swoop you will be helping very many fields of science, and you will also be helping the supercomputer industry, but not directly.

By stimulating software to be developed, you help the developers of supercomputers, but more importantly you help the users of supercomputers. If we do not do that, we will find that Europe and Japan will beat us.

Mr. ROLLWAGEN. If I may, I would add briefly to that because I agree with everything Sheryl said.

The additional comment I would make is that it is true that supercomputers and this technology act as a catalyst in all kinds of fields, probably every technical field, and furthermore it touches as opposed to the supercollider, for example, which is tremendously valuable, I think, in a rather narrow area of physics which is very important but still rather narrow.

This particular technology touches all fields of science, and furthermore, it touches all segments of society that are dealing with science, whether they are government laboratories or universities or businesses and particularly in the context that you are talking about in international competitiveness, it is in that business area where we can have the greatest impact.

I do not think it is a coincidence, for instance, as a small example in the automobile industry we have 16 of our systems installed. Only three of them are in this country. The rest are overseas and perhaps the more aggressive work is being done overseas, especially in Japan, and that shows up in the relative performance of the companies involved in that industry. And so it has gotten an awfully broad base to work from.

Senator BRYAN. That is an excellent lead-in to the next question I have. I am informed that in this country a majority of the supercomputers are owned by the government, the federal government, primarily. In Japan it is just the reverse, the majority is owned by the private sector as opposed to the public sector.

If I am correct in that information, what conclusions can we draw from that?

Mr. ROLLWAGEN. Well, I think, first of all, I believe you are correct. I am not as familiar with the gross data as I am with our own information and certainly based upon our information, that is absolutely correct although I must say in the United States now in terms of units, the industrial sector is sort of catching up with the government.

I think it reflects the fact that high on our national priorities is national security, particularly military national security, which has been a major driver for this and why there is so much activity on the government side.

On the Japanese side there's a tremendous priority on industrial and economic development and competitiveness and, therefore, an encouragement for the companies to get involved.

There's one other factor, though, which I think is the key one, and that has to do with macroeconomic factors such that the cost of capital in Japan is perhaps five times less than it is here in the United States, and these are expensive devices still and involve a lot of expense to support them.

And when you apply a 15 percent sort of threshold level, cost of capital to making an investment like this, it is very, very difficult to justify in advance. If you're applying a 2 or 3 percent cost of capital, you can be a lot more patient. You can take some chances and experiment, and I think that may be the key factor.

Senator BRYAN. Dr. Handler, would you agree?

Dr. HANDLER. Absolutely. I would even mention that one of the very large companies in Japan that has been speaking to us has been talking about wanting to make a very large financial commitment in this area. I asked the question how can you rationalize doing this. It was a very large commitment.

They responded that they had not made a profit in 15, 20 years, and were proud of it. And I said what? And they said, we have grabbed a big piece of the market share and that is more important to us.

And so it is true that Japanese companies can afford to invest heavily in supercomputers because they are not raked over the coals if they make an investment that does not pay off immediately.

And so it is hard to compare the U.S. government and corporate customer balance here versus overseas.

The issue is who can make the investment for the future, and in this country the government has always played the role of stimulating that kind of commitment.

Senator BRYAN. One last question before the light goes on, I understand the advantage of being able to share all of this massive amount of information that is generated and the need for linkage over a network.

What is not clear to me as a layman is why isn't that occurring in the absence of the piece of legislation that our chairman has sponsored here. What is it out there that is an impediment, a disincentive, if you will? Why isn't it happening without the kind of initiatives that the chairman and other members of the committee are working on?

Mr. **Rollwagen**. If I may, from my perspective, first, because we observed this, it is happening to a limited degree within specific organizations that companies and government laboratories have, in fact, set up high band with networks to use a resource that they own and make it available to their people.

They are able to justify that because they are large and they have the financial resources to do so, and they have been very successful and we can point that out.

The problem is, and I think Sheryl made this point earlier quite well, our country is based on a different program, kind of, a different way of doing things where smaller is better, organizations are smaller, they are diverse, they are spread out, and no one of them has the resource to put in the infrastructure by itself.

And I think also when Senator Gore alluded to the interstate highway system, it is the same thing. What trucking company could afford to put up an interstate highway system, and yet they all obviously benefit from it. I think that is the problem, a chicken and egg problem.

Senator **Bryan**. Thank you very much, Mr. Chairman.

Senator **Gore**. Very good.

Senator Lott.

Senator **Lott**. Thank you, Mr. Chairman, and thank you for your testimony. I thought it was very fascinating, and this whole field is fascinating, and I think to us as members of the Senate and to the American people, but we still have a very small understanding about what we are really talking about. You did a magnificent job of putting it in a very understandable language. And, Dr. Handler, I appreciated your comments, too.

I would like to pursue the question that you were just answering, referring to the legislation, and I do not mean to be putting this legislation down, because I think it may be something I would want to support and would want to do, but I have questions along the lines of why, why are you not doing it yourselves, and really should not this burden of networking really be in your industry, in the private sector? After all, you are making some pretty good money now. And I think you already responded to that question, and so I will not pursue it further unless you want to respond further at this time.

Mr. **Rollwagen**. No, and I will say that we are certainly prepared—can I say it this way—to do our part? Obviously it serves us well, and we want to contribute to that effort, but to do it alone, we cannot, speaking as a company.

Senator **Lott**. Keep in mind that when the government does things like this, whether it is a superhighway system or other areas, we do not always do it very well, and there are certain problems or controls that come along with it that you would not have necessarily to deal with if you did it in the private sector.

Well, the next question in that regard is security. Are there not certain problems in the area of sensitive information or security breaches that maybe we would not be able to control in this networking system?

Dr. **Handler**. There are networks in existence now. This is a network that would simply make more people able to use the network and it would be faster, so I do not think it changes the security issues.

They exist now. They are being dealt with. It is a serious issue, but it is a solvable issue.

Mr. ROLLWAGEN. I agree, and we already handle a lot of sensitive data, and there are safeguards that have to be put in place, and no system is perfect. There are going to be—there are going to be problems, no question about it.

Senator LOTT. My last question is one of management. Who would manage this network? How would we set that up to make sure that we are meeting the necessary requirements of facilities and services?

Senator GORE. Would the Senator yield on that? They may not be comfortable answering that because it is really a question directed at the legislation.

They are in favor of the network but may not be prepared to say who is going to manage it. Under the legislation that is pending it would be managed by NSF, which has had experience with the NSFNET.

The witnesses may have thoughts they wish to share on that, but I would just say also that we plan other hearings on the security question. I agree with the witness' statement, but I think as far as control under the legislation, it is NSF.

Senator LOTT. Well, I think it would be in your interest to take a look at the management, who that would be and how it would work, and to make sure that it is done properly. You may not be familiar with the specific legislation, but in general terms if you were setting it up, do you have any thoughts on what should be included in the management?

Mr. ROLLWAGEN. We would want to be a part of that process, of course, of deciding how it is to be managed.

We are very comfortable with the National Science Foundation. They already manage a large network—not of this capacity, and our systems are used to that network, and we realize they have the technical capabilities and the skills to deal with it, and we do not feel excluded from the process. We feel that we are consulted and can contribute to that effort.

Senator LOTT. Maybe I should ask you a more basic and elemental question. What would you, Cray, expect to get and to benefit from a network like this?

Mr. ROLLWAGEN. The primary benefit to us is additional users, additional people having access to the products that we make and the technology that we provide, and obviously there is economic benefit to that because presumably there would be more computers installed and purchased and so on, but that is really not it. I mean, that is important. I do not want to deny that, but let me put it this way.

At Cray Research now we consider our largest single technical resource. It is not Seymour Cray, it is not our engineers and scientists. It is not our laboratories.

It is our customers, the 170 or so people that are telling us every day how they are trying to use our machines, where they are running into trouble, what kinds of things we need to work on. That is our biggest technical asset, and the software that they developed to overcome some of those problems. That would be the primary benefit that would accrue to us that greater understanding of our own technology that comes from its practical use by scientists and engineers all over.

Dr. HANDLER. I would like to just add to that the number of people that are using supercomputers today are countable in the thousands, and that has to be in the millions if this country is really going to exploit this technology in an aggressive way. From our perspective, it is not clear to us whether we will sell fewer or more machines with the network. We could argue it both ways.

The reason we are interested in it is exactly what John said. It will clearly enable more people to use the most powerful supercomputers. It will not be that all of a sudden everybody is on it, and then that is the end of the problem. It will increasingly bring more and more people onto the supercomputer usage range. And that is so important that we are less concerned about whether we are going to sell either fewer or more machines.

Senator LOTT. Thank you very much, Mr. Chairman.

Senator GORE. Thank you very much.

Senator Robb.

Senator ROBB. Thank you, Mr. Chairman. I am fascinated by the potential implications for what supercomputers can do for today and for tomorrow. In addition to this legislation the government should be thinking about preparing for the potential revolution in virtually everything we do. Beyond the jurisdiction of this committee, beyond science generally, we have to be prepared to meet the needs of the late 20th-early 21st century and recognize the impact that supercomputers and everything linked to the exponential growth of technology and information will have.

Are there things that government ought to be doing in preparing for this revolution that go beyond the limited question that we are examining today? It seems to me that the governments of nations that prepare best for this revolution are going to be assisting the private sector to be much better positioned to be competitive in this global marketplace that we are talking about, where we go beyond that thousand customers and we really do have millions of individuals and companies using supercomputers.

Dr. HANDLER. I do not know if you can tell, but John and I are sitting here ready to pounce.

You have asked the magic question. The world is going to change and is America going to be in the position of leadership as we are changing? And that is the critical question.

I cannot stress enough that I think the network is a very important thing to do, but more important even is to raise as a nationally visible issue that America must be the leader in high performance computing. I think that having that goal clearly articulated is very important.

It is not a widely held goal here. It is understood and known in the supercomputer industry, but not beyond, and that is different than in Asia and in Europe. People there more commonly talk about needing to be, the leader, to leap ahead and become the high performance computing leader.

So, having the goal is important. And I think it is also important to recognize the structure of how America works best, which is all this diversity and to figure out how we can fuel that.

And as we know, in Japan these monolithic organizations that are very vertical and horizontal have the ability within their organizations to be both manufacturers and users and sort of the whole gamut of roles that are necessary, whether its subcomponents.

We do not have that here in singular organizations, and we should not. So, what we should do to take advantage of our strengths is to provide incentives and in particular I think tax incentives would be a very powerful catalyst if America provided a tax credit for all the companies and individuals that buy software for this new generation of supercomputing, you would see so much software developed so quickly that the industry would just blossom.

And this would not be a direct benefit to the supercomputer industry. It would be a direct benefit to the users of supercomputers. And if we encourage and provide that kind of leverage for the users that enables them to take a risk to go out and be the first users, the first exploiters, then we will do this on a massive scale, and I think it will be hard for anyone to compete with that.

Mr. ROLLWAGEN. If I may, Senator, I would even put it in—I agree complete with what Sheryl said, and I would put it in even broader terms. I think the world is changing. I think the world has already changed in some very fundamental ways and, in fact, I think the terms of trade have changed, or at least the objects of trade have changed, and there is no question about this in my mind in the nineties and in the next century that we are dealing in the information age, and we are trading knowledge, and this changes the rules of the game, and it is very difficult, it is very challenging for government to deal with that.

For example, a simple example, if we are dealing in knowledge—and I will take the simplest case—if I have knowledge that I am going to trade, that I am going to sell, I can sell it to you for money. And it is interesting because after I sell you my knowledge, I have your money and I also have my knowledge because I cannot exactly erase my mind.

And that sounds like a good deal for me, because I wind up with both. But you know what? You have my knowledge and you have your knowledge, and that is worth a heck of a lot more than any amount of money that you paid me.

And so what we are talking about is entering a realm where we are not in an exchange or bartering economy. We are in a sharing economy, and the only way to make it is to create value through the creation of new knowledge and new understanding of the universe, so to speak.

And the best way to increase my understanding of the universe is to trade everything I know for everything you know. And I wind up having your knowledge and my knowledge, and that gives me another lead forward in my understanding of the universe.

But it also shuffles the deck, and—so we are starting from the same place over and over and over again. And if we are going to be competitive in the long run, we have to be prepared to do that, and have the confidence to live by our wits.

Senator ROBB. Are there any structural elements of society today, either in government or in the private sector structure, that are inhibiting our ability to move rapidly toward that long-term or short-term objective? Is this growth really going to occur at the rate that everyone seems to predict that it will?

Mr. ROLLWAGEN. I will give you one specific case and maybe it is too small, and I am sure Sheryl can jump on that, too. There is a whole question of technology transfer ranging from the FSX to the getting of export licenses for instance, in our technology, and there is a great tendency because it is important, valuable technology to keep it to ourselves and not to share it or perhaps to take that technology and sell it for money or market share. Both of those I think are the wrong answer, and we have demonstrated that clearly time after time.

I think what we have to become is more expansive. There are national security issues involved with exchanging technology that are very important and need to be dealt with, but in that case I think we do tend to slow ourselves down in our own development of knowledge by being too restrictive in those areas.

Dr. HANDLER. Could I comment on that because that is such an important issue to us and it is probably the single biggest thwart to our expanding as rapidly as we could overseas. We do have a very aggressive and adventurous new technology, there has been some concern about whether this should be exploited.

Now, the interesting thing is foreign countries are developing hardware that is very similar to ours. It is very easy once we have a product on the market to—for someone else not only to copy it, but to do it better. We have seen this happen over and over again.

Holding our hardware back does not really protect America. There has been a lot of thought about having us demand something back from foreign countries when they buy our hardware so that they give us something in return.

And I think the point that John is making—and if he is not, please correct me—but this is certainly a point I am making—is that when we sell our hardware overseas without any barriers, we get access to customers and learn how they are using the machine. We use that to develop our next generation machines. There is nothing that is more important to our future than being the most knowledgeable, having the deepest understanding of how our products are being used so we can fold that back into successive generations.

If we only have U.S. customer user information and do not use foreign user information, we are going to fail. We do not have to ask anything back in exchange for selling our machines. All we have to do is get our products in the hands of customers overseas and learn from them.

We work closely with customers, most computer companies do. I think supercomputer companies do even more than others because we often deliver support people with the product. They live at the customer's site. Every one of our computers is delivered in a box, and there is also a person who is not in the box.

So, getting our machines blanketed everywhere so that Japan and Europe develop their software on John's and our technology on our hardware, that is what we need to do, and we cannot do it fast enough.

Our companies want to do it fast, but there are barriers to it.

Senator ROBB. Thank you. I could go on. I have exceeded both my five minutes and the jurisdiction of this committee, but I appreciate your answer.

Senator GORE. You would be surprised about the latest interpretations of jurisdiction.

Let me just interject one comment, because you and I have had discussions over the years on government's role, and I think one of the things that there has historically been agreement on in the United States between people on all points of the political spectrum or the ideological spectrum is on infrastructure because it lifts the ability of everybody to compete and be productive without picking winners or losers per se. Would both of you agree that this network is the—this is a dangerous question because I do not know the answer—but would both of you agree that the creation of this nationwide network is the single most important thing we could do to boost the productive use of computing capacity in the United States?

Mr. ROLLWAGEN. I will answer the question yes, but let me modify it. It is perhaps the single most important thing we could do. And it is clearly a necessary condition, but I am not sure it is sufficient. I think that the rest of the accoutrements need to come with it in terms of the software, the education and training. By itself, it will not do it.

Dr. HANDLER. That is precisely my point. I think the difference between the highway network and the computer network is that, on a highway, people can go out and see for themselves what the opportunities are.

On a computer network, it is much more abstract. And so we have to do other things. We have to put the infrastructure in place, and we have to provide other incentives so that the computer network is used productively, and not primarily for E-mail.

Senator GORE. I apologize to Senator Pressler for interjecting that question.

Senator PRESSLER. That is all right.

Senator GORE. Please go ahead.

Senator PRESSLER. My question might be directed to the Chairman, or to our witnesses.

I note that the bill would tie together universities and databases. There is one database I am very interested in, and that is the EROS Data Center near Sioux Falls, South Dakota, which archives the LANDSAT pictures, and it will soon have a NASA flag flying there, because the warming of the earth issue, the pictures of the earth—we have an antenna there now that is capable of taking pictures down from a spot in Mars, and some of the foreign satellites. Pictures of the earth, basically.

And with the current interest in the warming of the earth, and the various studies that are going on, these archives will become like the library at Alexandria, as my colleague has said.

But can these be processed in supercomputers, and what will be the advantage to the EROS Data Center to be a part of this network?

Dr. **Handler**. Yes. In fact, it is my understanding that only 10 percent of the LANDSAT photos that are in existence have been even looked at—have been analyzed. And so, here we have 90 percent of the results of our efforts are being ignored. And, in fact, supercomputers can do just that. We could have all of those in computer readable form that would enable anyone that has the sophistication to see the forest through the trees.

It is interesting that we spend many millions of dollars studying about the greenhouse effect, and the ozone problems. In fact, you can look at those photos and start to see those problems a decade earlier, rather than commissioning all those millions of dollars in studies after the fact.

Senator **Pressler**. Yes. That is music to my ears, because you are exactly right. We have the pictures in Sioux Falls, South Dakota, that nobody is looking at.

Dr. **Handler**. Exactly.

Senator **Pressler**. And we will send them to anybody. But now, how could supercomputers fit into this process?

Dr. **Handler**. Well, in particular——

Senator **Pressler**. Will we need more people at Sioux Falls to do the work?

Mr. **Rollwagen**. Of course, Senator. Yes.

You do, and I think Larry Smarr from the University of Illinois, the professor there that runs the supercomputer center, has a nice way of saying it. He says, "We invest millions, hundreds of millions of dollars in steel, iron and steel, and very, very little in silicon." And it is the silicon that provides the intelligence to the iron and steel, to the satellites, to the data-gathering systems that we have.

The fact is, the supercomputer, whether it is from Sheryl's company or mine, can, in fact, comprehend, if I can put it that way, the data that you have there.

The Senator was talking about the encyclopedia every second. Those pictures contain, actually, much more data, and much information than that. It can be transmitted to a system. The system can call that data, can detect patterns, can then display those patterns for a human being to analyze, so the human being does not have to go through all of the data. It may still be that no human looks at more than 10 percent of the pictures, but he is going to look at the right 10 percent.

Senator **Gore**. Would the Senator yield on this point?

Senator **Pressler**. Of course.

Senator **Gore**. Part of the legislation also calls for a national digital library initiative to develop infrastructure that could make it easier to get useful information that could then be transmitted across the network.

Just to use the example that you cite. If the network was in place, if the switches and the software were quickly developed, then I could see an atlas company, the National Geographic, or whatever, making regular use of digitized pictures from LANDSAT, sent from Sioux Falls over that network.

We could create a photographically correct globe of the world with supercomputers supplying the available topographic information at the same time, and create with the cheap materials, a photographically, topographically, correct globe of the world that would allow students to be their own satellite photographers with magnifying glasses. Actually, I have asked NASA if they would be interested in this, and they are.

Without supercomputers, that would be an unthinkably difficult task. But since all the information at Sioux Falls is stored in digital form—with this kind of infrastructure, we could make the data available for that application very, very quickly. And there are many other applications that we cannot even imagine.

Dr. HANDLER. Can I just address two points of that? One is the ability to have access and see patterns in amounts of data that are just too large to fathom.

And the other is doing it in real time. And an example of why that is important—a question was asked earlier about whether Japanese companies, or the government buys more supercomputers. And John had an approximate answer, but he was not sure, because he only had his own data.

If we had the network and huge supercomputers available, and the Library of Congress were on line—which will happen some day, we could have—while we were sitting here, asked an English question, using normal language, saying exactly the question that the Senator asked, and you would have your answer.

Now, the reason that is important is, getting your answer now while you are thinking about it makes a difference. If you get that answer a week later, when John looks it up and sends it to you, you might not even read it.

But being able to get what you want when you want it, whether you are a scientist doing experiments, or a businessman making a decision, or a legislator trying to make good policy, having access to information in a timely way is the difference between being the leader, or being a follower.

Mr. ROLLWAGEN. I bet they think we are real people sitting here, instead of simulations, right?

Senator PRESSLER. Well, I think it is——

Senator GORE. A very depressing problem to consider.

Senator PRESSLER. I will just conclude by saying that I think it is a very exciting thing that as we enter the process of making appropriations for studies of improving our environment, and anticipating what is happening, and the warming of the earth, and so forth, not necessarily whether it is at Sioux Falls or not, but wherever it is, it is important that these databases are in a computer someplace where somebody can access them quickly.

And I am told that we have a lot of information available that could be computerized, but that it takes so long to process it, and it is so clumsy to get at it.

And I am wondering, Mr. Chairman, if staff could prepare a, sort of a one-page paper in this area. Some of these pictures that are going to be used in the EOS Program, in the study of the earth that we are launching, and which NASA is going to conduct. How would that tie in—how would the databases be available to the common citizens through the supercomputers? Maybe that has already been done.

Senator GORE. We will jointly instruct the staff to do that, and let me go one step further and suggest that you and I jointly prepare a Pressler Amendment that specifically addresses the need for improving access to the extremely valuable digitized data at Sioux Falls to make it readily available as part of this network.

I believe that is imminently in the national interest, and we will certainly do so.

Incidentally, on the Mission to Earth Program, the single biggest challenge by far is not orbiting the satellites, or building the instruments, or anything else—it is handling the data flows that are going to be coming down from earth to orbit, which will be many, many, many orders of magnitude, larger than anything we have dealt with before in the earth sciences.

Did you want to comment?

Dr. HANDLER. I just want to—just remember, when you buy a jumbo jet, you really are buying more paper weight than the metal weight. That is very significant. Our society is changing, and if we do not have access to this information in a very powerful way, and before others do, then we will lose our lead in so many different areas.

Senator GORE. Is it true that your machines can, when applied to satellite observations, detect the growing of leaves?

Dr. HANDLER. There is no reason they could not.

Senator GORE. I was told by Danny Hillis. Is he here with you?

Dr. HANDLER. Is that what he told you? Well, then it must be true.

Senator PRESSLER. The pictures do. You can tell exactly what stage the crop is in, if there is a drought, if there are bugs in it, the whole thing.

Senator GORE. Well, we have kept you all—can I ask just one other question? I talked about this information infrastructure in terms of the network digital libraries—are there any other items that you would recommend be a part of that infrastructure?

Mr. ROLLWAGEN. I do not know if this counts, but one of the fascinating things to me is that there are two reasons to sort of centralize computer power—have it in a single place with a powerful computer. One is access to data, which you are talking about. And the other is access to massive amount of computation.

And what is fascinating in both cases to me is that if it is centrally located—if people are accessing it on the network, then suddenly they are working together. They are cooperating on the same experiment.

I do not know if that fits in the same category, but a person in California can be working instantaneously with a person in New York——

Senator GORE. Right. They are in what they call a co-laboratory.

Mr. ROLLWAGEN. Exactly. And as I say, I do not know if that is database sharing, but it is certainly knowledge sharing and it is accelerating the whole process.

Dr. HANDLER. I would like to make another comment. When I talk about supercomputers in the U.S., I have to often explain what they are. When I talk about supercomputers overseas, I less often have to do that.

And a typical thing that I often hear is, a physicist will tell me, I am a physicist, I do not want to have to learn to use computers. And I find that less true overseas.

The significance of that is, we have got to do one of two things. Either we have got to change the minds of our business people and scientists, and make them programmers, which I do not think we should do, and do not think we could do. Or, we have got to provide an incentive—lots of incentives to getting software written, so that the users will use these machines.

Overseas, people seem to be less concerned about having to learn how to be technically sophisticated in programming. And we have to do something to compensate with that here.

Senator GORE. Well, with this network, we will see the equivalent of interchanges, where software designers can cluster and download their products into the network with very ready access to a nationwide, continent-sized market.

We have kept you all too long, and there are no discourtesies intended to our other witnesses. They are extremely distinguished, and we better move on.

Thank you all. Thank you both very much for getting us off to such a good start.

Dr. HANDLER. Thank you.

Mr. ROLLWAGEN. Thank you.

Senator GORE. Dr. Craig Fields, please come to the witness table. Dr. William Wulf, please join Dr. Fields. Together you two make up our second panel. And so I will extend a word of thanks to the members of our third panel, who are going to be called on to demonstrate a little more patience, but thank you all very much.

Dr. Fields, of course, is the Director of the Defense Advanced Research Projects Agency, based in Arlington. He is well known to this committee, and well respected by this committee.

Dr. William Wulf, likewise, is well known to the committee and respected. He is the Assistant Director for Computer and Information Science and Engineering at the National Science Foundation.

Without objection, your prepared remarks will be included in full in the record. We invite you to proceed as you see fit.

Dr. Fields, we are going to start with you. I want to thank you both for being here, and we will get underway with your statement first, Dr. Fields.

STATEMENT OF DR. CRAIG FIELDS, DIRECTOR, DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

Dr. FIELDS. Senator, thank you very much. I wanted to start by discussing a little bit about the role of the government in development of advanced computing in this country.

You are suggesting in your bill a rather active role for the government—a more active role, and people will always question, is that a sensible thing, is that a good thing?

One way to approach it is not from a base of speculation, but from a base of history. So I thought I would look backward for just a couple of minutes.

I am going to do this in a rather egocentric manner, because that is where my knowledge base is. I hope you will not mind.

DARPA began in the field of computing in 1961. And at that time, we gave our very first contracts to develop some new-fangled, crazy idea called timesharing, which I am sure you have heard of, and has taken off. And since then we have been investing rather heavily to distinguish the way we invest from the way the National Science Foundation invests.

While we support universities and national labs and companies, really most of it is investment in industry, and mostly it is an investment that is aimed at developing an industrial base of products in which the Defense Department and others can buy things.

And so we are concerned about research. We are concerned about basic research. But to me the mark of success is when, at the end of our program, you can buy some new product—some new thing. And it is also worth noting that about 60 percent of our contracts in this area go to side adventures, and other small businesses.

Now, in operating in this way, what has happened? Well, I already mentioned timesharing. I could talk about computer graphics. I could talk about the ARPANET and the INTERNET multiprocessor supercomputers, like the connection machine.

In fact, going back to the 1960s, to ILIAC IV, which some people might claim is one of the earlier supercomputers.

The whole field of audiovisual intelligence, modern computer operating systems, multex, Berkley unix, the multex operating system, a lot of office automation technology, like local area networks, and major active text editors, work stations, the mouse on your MacIntosh, the computer language, ADA, new microprocessor designs like refinement of risk, microprocessors, the so-called list machine, and so on.

I could go on at great length with this list, and it is a much longer list. The fact is, the government has played a very, very powerful and important role in the development and support of the computer industry in this country.

Unless you think I am a biased witness, and I am, I refer to you a book by Kenneth Flam of the Brookings Institute, called Creating the Computer, in which he documents in, I think, a very nice manner, the strong role that the government has played in this whole area.

We can all have our own speculations and notions as to why the government has been so successful in aiding the computer industry. And I will just put mine on the table.

Namely, that in the computer industry, we have a situation where the barrier that you have to overcome between an idea and a product, between an idea and success, does not require the massive capital investment that you need in other industries.

If you have an idea in the semiconductor industry, you may have to invest $.5 billion, or $1 billion before the first thing comes out of your factory. That is not the case in the software. It is not the case in computers.

And so this is an industry where, at least to date, the fact that we have had higher interest rates, has not hurt as much as it hurts us in some other industries which are very capital intensive. I am not sure that is going to continue, and I am not sure it is going to continue because, in fact, this industry rests on a base, and the base is the semiconductor industry.

To the extent that semiconductor industry in the United States is eroded, and we have become increasingly dependent on foreign-owned sources for those semiconductors. The enjoyments of the past may not extend into the future.

Now, with that said as background, then I think, rather—on substantial proof that government's involvement in the computer industry has been tremendously beneficial to the country, let me say a bit about the future.

I am investing about $.25 billion a year of my agency's money in computer technology, information processing technology, because I cannot think of any other single thing that is more important. To do it is absolutely critical for defense, and of course, defense is where we are centered.

Our investment now is really in three areas, although there are small investments in lots of areas. One is in high-performance computing. Not only the computers and the architecture, but the underlying microelectronics, the software and operating systems, and so on.

Secondly, its in the area of computer networking. And, in particular, very high performance, on robust networks. It is pretty easy to get high performance on reliable networks, or high, reliable, low performance networks. Getting both is not so easy.

And, thirdly, is the area of artificial intelligence. Two of those areas are central to your bill.

I am as dedicated as I can be to assuring that in those three areas, the United States is number one in the world. And our investment is really aimed at that, and we are continuously doing things about it.

I do not know if you saw the press release that the Defense Department put out on the 5th of June about our latest endeavors, in which we talk about our working relationship with the Department of Commerce, to assure that anti-dumping is enforced diligently, and intellectual property rights are respected. Something that is in their purview, working with them, dealing on the Committee on Foreign Investment in the United States.

We are starting a joint activity with Cray. You already know about our investment in Thinking Machines Corporation.

I am working with the Department of Energy in Lawrence Livermore Laboratory to try to transfer software that is applicable to the supercomputer business out of our university labs. And a variety of other things. So this is a continuing investment and a continuing endeavor.

Two of our projects have pithy names. One is called the Teraop Project, and as you can guess from the name, the goal is to make a base of technology and some actual machines that are going to allow you to do a trillion operations per second.

I am a little sorry we started that now, because it is seeming easier than it once did. Really, we should have started a Petaop project, 10 to the 15th operations per second.

Senator Gore. Well, how do you spell that?

Dr. Fields. P-E-T-A-O-P, because 1,000 times a Teraop is really a challenge. The Teraop is seeming to be within our grasp.

Senator Gore. I knew that.

Dr. Fields. And the other project is our gigabit networking project. And there, of course, we are talking about networks that can deal with a billion bytes per second or two, or three billion bytes per second.

The rationale for the Defense Department in being involved in this technology development is so well established that it does not need review. And despite your flexible views of jurisdiction of committees, I do not think you really want to go into it in detail here.

But our needs in anti-submarine warfare, and an analysis of data of various sorts, intelligence data, means that we absolutely have to have these machines. It is not optional.

Senator GORE. May I interject a brief comment there. I am a member of the Armed Services Subcommittee on Defense Industry and Technology, chaired by Senator Jeff Bingaman. I might say that he is personally quite familiar with the initiative that we have underway here, and I have been working with him to enact measures in the defense bill, which are complementary to this effort, and directly address some of the matters that you discussed. Some of them, of course, are highly classified, but we are going on that track as well.

Dr. FIELDS. Yes, I know, and I appreciate the support in dealing with broader issues that you are concerned with on this committee, I struggled over the last couple of days with a way to give you a sense of scale for what these improvements really mean, because you keep hearing of gigaflops, and Petaops, and so on, and it is hard to understand. In the case of computing speed, it is easier than in the case of networking speed.

I just did a simple calculation this morning, and basically, the difference between today's supercomputers and these Teraops computers, is the difference between getting an answer in an hour, and getting an answer in six weeks.

If you are going to have an advisor, I might be willing to wait an hour to get an answer. If I had an advisor that made me wait six weeks, I would probably would want to get a different advisor.

In the case of gigabit networking, as I say, it is a little more complicated, because you have to identify things you can do with it, that you cannot easily do without it.

Not knowing quite how to address that, I decided, rather than looking forward, I would try to look backward. And here is how I have done that, Senator.

As you know, we are talking about an improvement in networking speed that is about a factor of 50,000. What I decided to do was to look backward at the last improvement in networking that we had, which was of the size of a factor of 50,000, and saying, what did we get for that improvement, or stated a different way, had some governing body at the time, chosen not to do so, what we would not now have, if we had not purchased that factor of 50,000?

And you can do that calculation. And let me tell you some things we would not have.

We would not have FAX machines, or cellular phones, or radio telephone. We would not have mailgram, or electronic mail. We would not have the telephone, or TELEX, or telegram, or air mail. We would not have mail. We would not have pony express. In fact, if you go back to a factor of 50,000 backwards, we would have smoke signals, and we would use flashing lights with shiny things from the top of hills.

What I found surprising from doing this calculation is that the improvement in the band width of long-distance communication among people that we have achieved, sort of from the beginning of mankind to now, is that factor 50,000. And that same factor of 50,000 is what we are trying to achieve over the next few years, developing gigabit technology.

Three million years ago, if someone had said, what are you going to get with a factor of 50,000, I doubt if anyone would predict cellular phones and FAX machines. I hope you are going to excuse me if I am not imaginative enough to say exactly what is going to happen with our next factor of 50,000, but, obviously, it is a big issue. It is not a small issue.

Now, to more graphically present this. We have tried to assemble a live demonstration for you today, and let me tell you that I face this with some anxiety since I have never seen a live demonstration work.

[Laughter.]

On the other hand, we do have video tape backup which unfortunately uses imported equipment, but what we are going to try to do and let me just give thanks to Lawrence-Berkeley Labs, the University of Southern California Information Sciences Institute to Sun Microsystems, as well as to our staff internally for arranging this to show you what it would look like to a scientist seeing a calculation. This happens to be a physics simulation, seeing a calculation with today's networking speeds of about 54,000 bits per second.

The next step up, 1.5 million bits per second. The next step up about 45 megabits.

Senator GORE. Are we supposed to look at this screen?

Dr. FIELDS. I hope so.

Senator GORE. Can the audience see?

Dr. FIELDS. There is another screen back here. And then lastly, at about .25 gigabit, which will give you some sense of scale. So if we could start with the slowest one, which regrettably is pretty boring because it is so slow.

This is today's experience. We will just have to sit through it to get to tomorrow's experience. Actually is there anyway to speed this up and to simulate getting to tomorrow or do we have to wait?

Senator GORE. This is one flop, right?

Dr. FIELDS. Actually, since we are in communications, I think of that as a millabit per fortnight.

Senator GORE. That is all right. I think of my presidential campaign as a gigaflop.

Dr. FIELDS. That is the next step up in speed. You sort of wait a while and you get a picture.

This is 45 megabits. This is a simulation of what would happen when you get up to 45 megabits. Now you are starting to get some concept formation. You can get some impressions. You might want to proceed apace. And now you see what happens when you get to the higher speeds, still.

The point is that at the lower speed, you are seeing it sort of like trying to see Citizen Kane and tell you how the masterpiece by just seeing a frame at a time you are not going to do it. As you get to these higher speeds it is possible to form concepts, get understanding, and that is what you are trying to achieve with the network.

Senator **Gore**. What exactly is this? What is the picture? Does that matter?

Dr. **Fields**. Well, it actually does not matter, but this is a simulation that is sort of like a calculation you might use for engine design or explosive design. And if you want to go onto something else we can do that.

It is coming. Of course, there are medical applications as well. The point I am trying to make is that I did not want you to get the impression because it was Defense Department sponsorship. We were particularly fixated on flames and explosions and so on. There are any number of applications for this technology, and this kind of motion that you see, you just will not see at the lower band width. You will not form the concepts; you will not get the impressions you need this to generate science. You need this to generate technology and that is exactly your point.

And I think that is good lead in for the kind of remarks that Dr. Wulf has to make. Thank you.

Senator **Pressler**. What is exactly the point here? Is it the accelerated speed or what?

Dr. **Fields**. Yes. The point I tried to get across from the first series of demonstrations was that at today's low band width 54,000 bits per second, you have to wait a long time to see one picture. And you will never see any of the motion. You will never see the connections. You will never see the changes at the higher band widths, you see it, and we think that is an important aid to thinking.

Senator **Gore**. The pictures are more complex. You can see them in three dimensions. They can be moving over time and can convey the meaning of much more complex interrelationships.

Dr. **Fields**. That is right. You can see change. You could not see that at the beginning.

Senator **Gore**. All right. We will hold our questions until after Dr. Wulf's testimony.

Dr. Wulf, please proceed.

[The statement follows:]

TESTIMONY OF DR. CRAIG I. FIELDS, DIRECTOR OF DARPA

Mr. Chairman, I am delighted to be here today to present DARPA's perspective on plans for creating a U.S. National Research and Education Network (NREN) to link university, industry, and government researchers with advanced computing and networking technology, including defense sponsorship of R&D which has led to some of the highest performance supercomputers in the world. We were also responsible for creation of the ARPANET and Internet, which will be the starting point for development of the NREN. We anticipate an active DARPA role in developing the multi-gigabit networking technology the NREN will require.

As other witnesses will be discussing the importance of the NREN, I plan to confine my remarks to the issues which concern DARPA most: the process of developing multi-gigabit networks, and the relationship of the network project to other high performance computing activities.

GIGABITS NETWORKING RESEARCH AND DEVELOPMENT

DARPA initiated a program in networking at speeds of billions of bits (gigabits) per second in fiscal year 1988. We began with a study that showed conclusively that this technology is ripe for development and can support many important defense applications. While our program was intended primarily for National Security purposes, networking also represents a means of increasing the effectiveness of our research community through remote access to powerful resources and new capabilities for collaboration. Based on that study we have planned a program that will support deployment of gigabits networking that will be sufficient to create a multi-gigabit NREN sometime after the year 2000.

DARPA's focus is on building networks using communications to link computers, by switching the data among the links. We are not emphasizing the communications links because many of the required communications facilities are already available. Nevertheless, much research is still needed to achieve a gigabit network. Switching systems must be developed to handle the type and volume of data involved; algorithms must be designed to manage network resources that operate at very high speeds; interfaces and operating systems must be built to enable computers to fully utilize the capacity that will be made available; fiber optics trunks and satellite links must be configured to support the network; and all of these elements must be meshed into an integrated, coherent technology base. The result will be an important new capability for the military to transfer large amounts of data (such as high resolution images) between sensors, weapons platforms, computers, and command centers, within very short time intervals.

The same gigabits networking technology will apply to the Stage 3 NREN, providing a different set of benefits. Other witnesses undoubtedly will describe in detail the advantages to the research community of remote access and collaboration; equally important to the nation, I believe, are the benefits to U.S. industry. I believe that it is not overstating the case to say that a national network will be as important to industry in coming years as highways and railroads, mail, and the telephone have been in the past. Our intention is to work for early transition is that the gigabits networking technology will become available as a unique nationwide resource. For example, high speed networking will allow industry to ~~shorten~~ production cycles by coupling research, development, production, and inventory between various producer locations and among producers and users of manufactured goods. Ultimately, I would expect that consumers would be connected to this sort of network also. For example, a ubiquitous high-speed network appears to be the ideal medium for distribution of high-definition television signals.

Before leaving the topic of the NREN, I would like to comment on the question of the organization which will bring it into being. I feel it is important that an interagency group manage the development of the NREN. The reason is that, if DARPA and the other mission agencies are to rely on the NREN to support their researchers, they will require a voice in its governance. I view this point as critical.

HIGH PERFORMANCE COMPUTING PLANNING

Progress in high performance computing will require balanced efforts in high performance computing technology, systems and applications software, and basic research, in addition to the national network. To provide the research, development, and manufacturing environment America needs in the next century, it is essential that all of these areas advance together. In this regard the position articulated in the Office of Science and Technology Policy's document A *Research and Development Strategy for High Performance Computing*, prepared with the assistance of the Federal Coordinating Council on Science, Engineering, and Technology (FCCSET) provides a good basis for planning. It is extremely important that such a strategy be pursued.

In one of these areas, high performance computing technology, DARPA currently has a program in parallel computing that is widely acknowledged to be the strongest in the world. We have a proven track record in stimulating development of this technology and managing the major prototyping efforts required to demonstrate its effectiveness. Our current goal is to create technology for general purpose parallel computing systems capable of scaling up to trillions of operations per second - "teraops" - and we believe everything indicates that we are on the right track. By way of comparison, the fastest computing systems today are capable of a few billion operations per second.

We have been cooperating closely with other Federal agencies in the FCCSET process, with the result that there is a consensus among the principal players that a two-pronged attack is needed to maintain U.S. leadership in supercomputing. Other agencies are continuing to stimulate the commercial supercomputing industry as they have in the past by purchasing early systems when new models become available to support their mission needs. At the same time DARPA is funding development of parallel computing technology and sponsoring prototypes of the promising parallel systems, to be evaluated in applications in cooperation with the other agencies and scaled up to larger capacities if warranted by evaluation results.

High performance computing technology is critical because the U.S. requires access to this technology for national security. Unfortunately we are losing ground in this area. DARPA is convinced that parallel computing is the high performance technology of the future.

Mr. Chairman, this completes my testimony. i have brought along some demonstrations that show some of the advantages of high performance computing and gigabit networks to the research community. I will be happy to show these to you if you are interested, and to answer any questions you may have.

STATEMENT OF DR. WILLIAM WULF, ASSISTANT DIRECTOR, COMPUTER AND INFORMATION SCIENCE AND ENGINEERING, NATIONAL SCIENCE FOUNDATION

Dr. Wulf. Thank you, Senator. I appreciate the opportunity to be here as well as your continuing interest in this important area. You introduced me as the Assistant Director of the National Science Foundation. I should tell you that in real life I am an AT&T Professor of Engineering and Applied Science at the University of Virginia.

I thought I would take this opportunity in my oral testimony to say something about what it is like to be a user of the research network from the perspective of a user of the network, what is important. Most of my academic life I was supported by DARPA and was a user of the ARPANET for almost 20 years.

I will try to focus on just four points.

The first one is that the overwhelming feeling out in the community is "Let us get on with it. It is time to move." The value of the national network is not a question. It is certainly true that some sciences and some industries have moved faster to develop a dependence on the networking technology, but it will become absolutely essential to most.

We have talked a lot about the supercomputers. That is not the only application. For example, the long-term ecological sites funded by the National Science Foundation simply could not operate without interconnection and communication by the network. In questioning a number of scientists about the use of the network, I was amused by the response that we got over the network from the director of NRAO, the National Radio Astronomy Observatory. And I am not sure I can quote it exactly, but it was to the effect that without the national network "life as they know it would cease." That is they could not do their science. The fact that the traffic on the network is growing at the rate of 20 to 40 percent per month, which is phenomenonal—I wish my investments would do that—is some indication, I think, of the demand.

Senator Gore. Wait a minute. How fast is the growth rate?

Dr. Wulf. 20 to 40 percent per month. It is doubling roughly every five or six months. As you know, we switched to 1.5 megabit backbone just about a year ago. That was a 30-fold increase in capacity when we made that change. Yet it is unlikely that that will not be adequate capacity in the next year. It will saturate.

So the value of the net is not a question. I think the management and technology plans are not an issue. They have been developed by the various agencies working together and have been vetted with not only the agencies, but the congressional staff, industry, academia—everybody that we could get to comment. Again, they are not an issue.

The issue is just getting on with it. Getting it done. And frankly, that translates into resources. The requests for those resources are working their way through the budget process. Everybody should understand that this is going to require incremental dollars. I know you understand that. We have made remarkable progress because of the interagency cooperation over the last several years.

But that is not going to be enough. We are going to have to move out.

The second point I want to make is that industry-government-university cooperation in developing the network is absolutely essential. I have heard some suggest that we ought to let the private sector do it all. And it is the "all" in that suggestion that is the problem. Dr. Handler and John Rollwagen gave you one reason why it was appropriate for the government to be involved. But I think there are several others that merit mention.

First of all, the ARPANET has existed for roughly 20 years. Industry has not come in to take over that function because the market has not been large enough. Part of what we need to do is develop that market. It is true that there will be experimental commercial municipal area networks next year. But they are experimental. They will cover only two cities. We cannot wait. I go back to my point: let us get on with it. By contrast, NSFNET right now connects over 250 institutions in 48 states. About 100,000 computers are accessible over the network.

Senator GORE. Do you have large waiting times now to get on the net?

Dr. WULF. For an individual institution? Not that I am aware of, no.

Senator GORE. For researchers? I mean I am told that some researchers have to wait, that they have trouble getting access.

Dr. WULF. There may be some areas which are not serviced as well as other areas. That is certainly the case.

Going back to my point about whether the private sector should do it all, part of the difficulty stems from a misunderstanding of what the word network means. It is certainly not just the physical infrastructure. In fact, sense, it is not even the physical infrastructure. The physical infrastructure is already supplied by the private sector. The government simply leases those circuits. It includes the basic packet, delivery mechanism. It includes services, most of which are supplied by computer companies and not by telecommunication companies.

It includes a great deal of support for what I refer to as affinity networks, that is, groups of people who share software, databases, bulletin boards and so on. And most of that is supplied by universities. The whole point is that expertise from all of those sectors, public, private and academic, are critical to making the network happen.

A centerpiece of the FCCSET implementation plan, which, as I am sure you know, is complete and is working its way through the administration, is the transition to commercial service. It is not exactly clear how that is going to play out. We are funding studies to see what the best way to do that is. But a centerpiece is the goal to make that transition and to do it as rapidly as possible.

Senator GORE. Let me just say that I certainly agree with that philosophical approach. And I do not see any obviously impediment to accomplishing that transition. There is an entry barrier now for the private sector that is a kind of chicken and egg problem. The demand is not there. The capacity is not there. Which comes first?

And if we simply rely on the market forces generated by voice communications and the low volumes of data generated by current commercial computer traffic, then the demand is never going to—well, sometime it would—but it is not soon going to reach the critical stage when a lot of investment will be pulled into the market place to create this privately. It is just not going to happen. By contrast, if we get over that hump with a government initiative, then it should be relatively easy to design a transition back into the private sector.

And we should eventually consider some kind of user fees that completely fund the operations. I mean I would think that that would be very, very easy to find the equivalent of a gasoline tax that replenishes interstate highway fund.

Dr. WULF. I absolutely agree with you. All of the agencies working together have carefully structured our implementation plan to involve the private sector, to keep them fully informed, as John Rollwagen said, to have input to the whole process as we go along.

A third major point I want to make has to do with the dual role of the National Research Network. Building the net is not a simple issue. It really has two aspects to it. One is the set of research and education which uses the network and the other is research on the network, on the problem of networking, for which DARPA is primarily responsible. The two are not independent.

The way that we use the network must impact the design. Interactive visualization such as Craig was showing you a few moments ago is radically different from television video images. It places entirely different demands on the network. It has different holding times and so on.

Data fusion from geographically disparate seismic sensors is radically different from either visualization or telephone or electronic mail, for that matter. So we have to understand how the network is used in order to do a proper job of designing it. To me that underscores why it is so important for this whole process to be an industry-government-university cooperative venture.

This is not a simple administrative government network. This is not a simple extension of the current networks to higher speeds, because the patterns of use will be different. The expertise to build a network is not wholly located in government or industry or in academia. We have to work together.

I think that the FCCSET plan and, frankly, the NSF cooperative structure to manage the network, is designed to encourage and even to require that kind of cooperation.

I will make one short side comment here, having now just emphasized the dual role of both research on and using networks. The National Research Council, when they critiqued the original FCCSET plan, encouraged us to make sure that research on networking did not interfere with the productive use of the network by researchers. We are very mindful of that. And so, for example, ourselves and DARPA have put out a solicitation recently for something called the RIB, the Research Internet Backbone, which primarily exists at least in its early stages to supply completely separate circuits that we could use for research on networking and, therefore, not impact our users.

The fourth point I want to make really harks back to something you said in your opening remarks and which I so heartedly agree with, and that is the need for balance in this program. The original

1987 FCCSET report called for a balanced program of high-performance computing, networking, software and algorithms, and basic research in human resources. And I do not feel I can overstate the case for the importance of that balance.

It is easy to get blinded by the shiny toys, by the hardware. But in point of fact, the algorithms, the software and the people who are able to use them are every bit as important to this entire process.

In summary, I think the most important two points that I wanted to make are the first and the last. Let us get on with it and let us keep it balanced.

Thank you.

[The statement follows:]

Testimony

Dr. William Wulf
Assistant Director
Computer and Information Science and Engineering
National Science Foundation

Introduction

My name is William Wulf. I am the NSF Assistant Director for Computer and Information Science and Engineering (CISE). I am on leave from the University of Virginia, where I hold the AT&T Chair in Engineering and Applied Science. Prior to that, I was Chief Executive Officer of Tartan Laboratories, a software company I helped found, and before that I was Professor of Computer Science at Carnegie-Mellon University for.

I am here today in several capacities: as the NSF Assistant Director responsible for networking activities, the national supercomputer centers, and computing research; as a member of the OSTP FCCSET Committee on Computer Research and Applications, and chair of its Networking Subcommittee; and as an interested scientist with experience with high performance computing and networking.

In response to an inquiry from this committee, The OSTP FCCSET report, "A Research and Development Strategy for High Performance Computing," (November, 1987), prepared with the assistance of the Committee on Computer Research and Applications, defined a balanced program leading to maintain U.S. leadership in high performance computing. It outlined four key components: (1) high performance computing resources (both supercomputers and local computing), (2) a national research network, (3) algorithms and software technology, and (4) a strong program in basic research and human resource development. During my tenure at NSF I have been working with the other agencies represented on the above FCCSET Committee to develop an interagency plan for the implementation of the goals laid out in this visionary report. This implementation plan, developed under the leadership of OSTP, has been completed and is presently under formal review within the administration.

The 1987 OSTP Strategy report, the current OSTP Implementation Plan, and the "National High-Performance Technology Act of 1989", share a vision of the role that information communication and processing will play in our future, and of the need to enhance the national information technology research infrastructure. They also recognize the investment needed to realize the potential and to avoid the erosion of our technology base relative to that of our international partners and competitors. Thus, I would like to focus on overriding issues of this shared vision, including several challenges that must be addressed in order to move forward.

I want also to emphasize the need for a balanced program of investment. In the excitement over the hardware, software and people are too often overlooked. Our experience with the national supercomputer access program proves that it is inefficient to ignore broader system and software issues. Several successful research projects have shown that with improved scientific algorithms and by "tuning" the software to match particular machine configurations, there is the potential to increase the performance of the machines currently at the NSF Centers by up to a factor of 20. This translates directly to improvements in capacity (ie., more access) or capability (ie., better modeling).

With the next generation of parallel supercomputers, early investment in basic software research and training people in the new technology will be even more critical to the overall efficiency of the program, and hence to the quality of the science which emerges. The same is true of the network. Advances in switching hardware, for example, are necessary to meet future traffic requirements. Yet it is the advanced information processing services, the automated databases, digital libraries, and multimedia communications that will make the network most useful to the researcher and to the industry.

The Present Status

The agencies represented on the FCCSET Committee recognize the importance of information processing and communication to research productivity, and consequently to our national competitiveness, security and prestige. At NSF, for example, that recognition has been translated into action. We have steadily increased our support for networking, supercomputing, and basic research and human resources over the past several years, and expect to continue the trend within our overall approved budget totals. Let me give three examples of what that has produced.

First, with the cooperation of the DARPA, NASA, and DOE, NSFNET has become the core of a rapidly developing "internet" which ties together and extends many governmental, private and international research networks. Because of our unique mission, NSF has become the de facto lead agency for the implementation of the National Research Network. The new 1.5 megabit/second NSFNET backbone became operational last summer. Via the midlevel and campus networks, it services over 250 research institutions, and provides access to over 100,000 computers. It is working well and the research community is responding; in fact, over the last few months traffic on the net has increased 20-40% per month. To meet the demand we expect to expand the number of nodes and links on NSFNET this year and, subject to budget constraints, to begin to phase-in the 45 megabit/second "Phase 2" backbone in the later part of FY90, almost two years ahead of what was suggested in the 1987 FCCSET strategy. This has occurred because of the dedication of the program staffs of the agencies and the extraordinary participation of our industrial partners.

Second, subject to appropriation of the FY90 budget request, NSF will have increased its support to the National Supercomputer Centers by 60% between FY88 and FY90. The NSF Centers will be upgraded to the latest supercomputing systems available, and staffed to maintain their leadership role. With considerable state and industry support of the NSF Centers, and the emergence of the state and regional supercomputer centers, the supercomputing capacity available to U.S. academic researchers has increased over 80-fold since 1984. The special capabilities of this class of machine have been well established in a broad variety of research fields.

Third, with the aid of the NSF "CISE Institutional Infrastructure" program, the production of PhD Computer Scientists has nearly doubled, to over 600 per year, since 1985. While this falls far short of projected needs (estimated by some to be 1600+ per year), it has made a notable difference in the ability of universities and industry to recruit able researchers, developers and educators.

These are small steps, but critical ones, in reaching the vision outlined in the FCCSET report. It is also testimony to what can be accomplished among collaborating federal agencies, and through Federal, State, university and industry cooperation. These steps demonstrate that we have the mechanisms and the will to apply them wisely and efficaciously.

Prospects
These are exciting times in computer networking, high performance computing, and in information processing in general: improvements in capability are accelerating, and competition is increasing. To be sure, there are problems to be solved to realize the potential of these capabilities, but history indicates that the payoff to the nation will be substantial. There are a great many issues to discuss, but first I want to concentrate on just a few of the benefits of information processing as I see them.

It is clear that networks increase the productivity of researchers, by allowing them to collaborate with a larger set of colleagues, by giving them timely access to data collected at remote sites, and by giving them access to unique resources. More importantly however, networks allow research to be undertaken that simply couldn't be done otherwise. For example, they permit interdisciplinary collaboration between researchers that do not happen to be collocated. They permit access to data and facilities that would not be accessible otherwise. They speed the spread of ideas, broadening the scope of problems undertaken by the research community. They aid technology transfer by allowing direct and fast interaction between people, and between people and research literature. That is why it is so important that research organizations in industry have access to the national network.

Less obviously, networks can also effectively increase the pool of researchers by making it possible, for example, for the faculty of four-year and minority institutions to participate in research activities based elsewhere. Apart from simple equity, we now that to be competitive in the future, we will require more participation from those parts of the academic community that are at present under represented.

It is clear that higher performance computing enables scientific investigations that were previously impossible -- because, for example, the data were too voluminous (as with the human genome); the sample sizes were too small to perform experiments (say, a few tens or hundreds of atoms); or because experimental environments were too hostile to humans (as in a nuclear reactor or on the bottom of the ocean). Important basic research problems which are beyond the capability of our most powerful computers yield to the advances in computing technology over the next few decades.

More immediate payoffs will be obtained in other areas as well, where the computational requirements match those of basic science. Engineering design and manufacturing are being transformed by computing. Supercomputing is being transformed from an esoteric tool of basic research to an accessible, working component of industrial development. If we are to keep up industrially, we need to accelerate the transformation not only by increasing access to the machines and training more people, but also by investing in the research and development on interfaces and applications software

Networks and computers separately will have a substantial impact, but it is their synergistic interaction in a comprehensive information infrastructure, exploited by a educated cadre of scientists and engineers, that will have the most profound effects. We cannot quantitatively measure these effects, and have no way to predict their precise future implications. However, I believe that what will emerge from exploiting more capable, less expensive computing and communications technology, and vastly more information in computable form, will have profound effects on science, engineering, and society. One can see glimmerings of this future in the major successes research networking has already had, for example, in medical informatics (SUMEX-AIM and National Library of Medicine's Medline), exploitation of VLSI technology design and prototyping services (MOSIS), and the NSF supercomputing effort. I have personal experience with the design of Ada, the new DoD computer language for embedded systems, where the involved community was located literally throughout the world. The work simply could not have been done without ready access to the ARPAnet.

National R&D Partnership and Competitiveness

Increasing the effectiveness of researchers is important, but the payoffs are compounded when the results of research are applied. The information infrastructure has a critical role to play in moving ideas more rapidly to the private sector, in enabling faster development of "defect free" products, and in generally increasing the "intellectual stew" from which our industries extract the ideas that are ripe for commercialization. Paper and people are still important ways to move ideas, but increasingly, industrial researchers and product developers require access to the same electronic community as the academic researchers.

The two-way flow of ideas in this enlarged community not only benefits industrial developers, but also provides stimulating, realistic problems and critiques to academics. You need to understand that in computing and networking, private industry has been a strong partner with the government for a long time, and with excellent results. NSF's experience with a predecessor to NSFNET, called CSNET, illustrates the importance of connectivity to industry research laboratories. Industry has invested heavily in internal corporate networking and many paid to connect to CSNET at rates which helped subsidize the university participants. They believed it was worth the price. Similarly, industry has invested heavily as partners in NSF's computing and networking research activities. Apart from IBM and MCI financial participation in NSFNET, there has been substantial investment by Digital, Apollo, Apple, IBM, and SUN in the EXPRES project, which explores advanced documentation and collaboration technology.

Industry has invested heavily in the National Supercomputer Centers and the entire scientific computing effort. In addition to the supercomputer vendors, Cray, IBM and CDC, many manufacturers of product ranging from mini-supercomputers, workstations and communications devices to specialized scientific software, have worked with NSF and with the Centers to realize the success of the program. Numerous industries outside of computing are affiliated with the centers, in areas ranging from biotechnology and aerospace, to mining, materials manufacturing, industrial design and economic forecasting. Many other examples could be cited as well for the computing research programs of other agencies, especially DARPA, which has had a substantial role in research that helped spawn the industry in the U.S.

American educational institutions have a history of leadership in developing and using computing and information technology. They pioneered network-based computing innovations such as timesharing systems, research software packages, standard operating systems, automated text retrieval, and computer aided instruction. NSF has working closely with EDUCOM, NASULGUC, and other educational consortia to make NSFNET and the supercomputer centers a reality, and to insure that they serve the needs of teachers as well as researchers. We will continue to do so, because the need for improved access to educational resources parallels that for research resources, because of the paramount need to educate a cadre of computationally literate scientists and engineers in universities and industry, and because the implications for improving educational productivity and saving scarce resources are profound.

These successful partnerships result not simply from industrial understanding about the long term economic payoffs from research, or support needs of university researchers. They are a product of the historic, enlightened cooperative investment policies of American industry, government and universities.

Challenges

The potential benefits of an information infrastructure will not be realized without meeting a number of challenges. The U.S. is not alone, and perhaps is not even the most aggressive, in recognizing the benefits and addressing the challenges of information technology. There are several types of challenges.

First, there are intellectual challenges. Advances in information processing require innovations apart from the physical technology. These include software and algorithms, programming languages and software engineering techniques, artificial intelligence techniques, new computer architectures, system design, prototyping and verification tools, information services (including directories, access and security controls, digital libraries with automated retrieval and analysis utilities, and tools for collaboration), and network management tools. Research in these areas is as important and as intellectually challenging as any for advancing understanding and promoting technology transfer. The challenge is to bring together the many disciplines and their specialized tools to work together on the problems involved.

Several agencies have made a start at meeting these challenges together. Critical areas have been highlighted by joint research program announcements between DARPA and NSF, and by joint support of projects by DOE, NASA, NSF and DARPA. NASA, DOE, the service agencies and NSF have worked together to effectively use their resources in providing supercomputing and networking access via sharing and consulting assistance. DARPA has focused sharply on advanced networking research, and now leads the coordinated efforts of the larger research agencies.

Second, there are substantial organizational and managerial, challenges which make it difficult to generate the consensus needed for concerted and sustained action. The problem is real and is exacerbated by the fact, as pointed out in the NAS critique of the FCCSET report, "Toward a National Research Network", that much of the federal, state, industrial, and university planning must be done without the benefit of a firm specification. The challenge is to build the coordinating mechanisms for the creation and management of an advanced facility that has essential components at every level: campus, state, regional, and national; across multiple jurisdictions, with mixed public and private ownership of segments. The existence of the present internet is proof that it can be done with sufficient commitment, flexibility, and shared resources.

The agencies have devised a plan to maintain the momentum that has brought the internet to its present state as the world's most advanced general purpose computer communications network. The plan will steadily improve and extend the existing internet to match the expected growth of traffic, while simultaneously developing the technology that will allow the merger of the separate agengy networks into a single "National Research and Education Network". The agencies have agreed that NSF should lead the effort to establish this facility as part of the nation's basic research infrastructure.

Third, there are exciting technical issues to be resolved. The "Phase 1" and "Phase 2" networks envisioned in the FCCSET report are reachable using today's advanced technology. However, advances in switching technology, communications protocols, routing, and accounting algorithms are necessary to realize further progress. Similarly, developments now underway at established and "start up" companies will lead to improved high performance computers. However, advances in basic semiconductor and optical devices, computer architecture, algorithms, and software will be necessary. A well coordinated, balanced basic research program by government, industry, and universities will be critical to achieving theses advances.

Fourth, of course, there are financial challenges. Mechanisms must be created to sustain the mix of research and service facilities that are the information infrastructure, to deal with the costing and accounting, recovery and subsidy problems of a system with central and user controlled segments. These might seem overwhelming at first glance, but we are dealing with a national resource which will be far less costly if developed in a coordinated way than would be the case if the parties at interest continue on separate paths. Only a part of the effort will be undertaken if the work proceeds as "business as usual," and time horizons will be stretched beyond what may be useful in terms of national competitiveness.

Fifth, there are the challenges of "transitioning the network to commercial" service." Note, however, that the word "commercialization" is used differently by various interested parties; in a sense a great deal of the present internet is already commercial. NSFNET, for example, is managed by a contractor via a cooperative agreement. The contractor purchases bandwidth from the telecommunications industry. It builds no plant of its own and will have no need to do so in the future. The contractor provides "backbone" services to link midlevel networks, and some specialized facilities, such as the NSF Supercomputer Centers. The midlevel networks are a mixed group characterized at present by some private, some public support.

For the most part the midlevels are themselves "private" entities. But they are not the telephone company or even like the specialized common carriers that have developed over the past decade. Their target "market" is so circumscribed that it is not attractive to the real firms in those business at present. That may change in the future, and the network configuration or the shape of the midlevel networks will have to be open to change to meet that eventuality. The campus networks that in effect provide the "final mile" and "customer premises" segments are similarly "private" entities (whether wholely owned by an institution or rented from commercial vendors). This is the model on which the internet is built: it is a cooperative venture which matches invention to necessity. That is the source of its flexibility to innovate and extend service and its capability despite the absence of a sustainable free standing market for its services.

The critical issue is one of policy. The goal of our effort is to move the resource as much out of the public sector as possible as soon as that is practical -- that is as soon as the users are able to purchase the capability and connectivity they require. We believe that the market will grow due to this effort, much as the supercomputer market has grown as more researchers have been exposed to the benefits of that kind of capability, and will intersect other developments underway in the communications industry sometime in the next five to ten years. By that time the federal role will have changed substantially.

Finally, there is the challenge of maintaining the balanced program of computing, networking, software, and human resource development. The information processing infrastructure is not just physical devices composed of glass fibers and silicon chips, and the worldwide competitive information industries are not exclusively manufacturers. The physical aspects of the infrastructure are easier to point to, and to touch, but without software to activate the hardware, and trained people to use it, the technologies are literally meaningless.

Conclusion

While I agree with the goals and directions of S. 1067, and applaud its recognition of an important national issue, I have a number of concerns about the bill. The bill has much in common with the draft OSTP implementation plan. Both are based on the 1987 OSTP Report, "A Research and Development Strategy for High Performance Computing". However, the bill seeks to accomplish legislatively what is more readily, over the long term, accomplished by the administrative actions such as those called for in the OSTP draft implementation plan. Legislating dates, processes and management structures will restrict an ability to be responsive to the future requirements for high performance computing. The bill's detailed legislative prescriptions of the OSTP/agency roles would, over a period of time, hamper the effectiveness of the interagency process as requirements and technologies change.

The following aspects of the bill are objectionable:

1. The bill authorizes appropriations to OSTP to be distributed by OSTP to various agencies. This is at variance with the charter for OSTP, which does not authorize such disbursement of funds.

2. The authorizations are incremental to an existing research effort (not specified in the legislative base) and cannot be uniquely related to existing budgetary categories.

3. Section 202 calls for an operation and policy making roll for FCCSET, in distinct contrast to its current advisory role within the 1976 OSTP charter. This new role cannot be readily incorporated within OSTP as an agency within the Executive Office of the President.

4. The authorization of additional, unrequested funding for NSF could have an adverse impact on NSF's programs and priorities.

While we support the objectives of the legislation, because of the issues that I have identified and several other legal, trade, procurement, and intellectual property problems in the bill, we must oppose S. 1067.

The near term payoffs of the new information infrastructure for research and education are large, but the longer term impacts are *far,* larger. We are privileged to be involved in the development of this infrastructure. A century hence, this era will appear a milestone to human progress. We are developing tools that embody our knowledge and amplify our intellectual prowess. The industrial revolution, by contrast, was merely the culmination of experience with tools that amplify our physical prowess. The easy-to-foresee payoffs of the effort today will be innovation important to R&D productivity and industrial competitiveness. Yet, since our intellectual capacity is the essence of what makes us human, tools to amplify that capacity cannot but have an impact beyond our ability to imagine.

Our severest challenges, then, are to support the visionaries, to catalyze the human resources to realize the potential of high performance computing and communications technologies; to accept the risks and commit resources to seed these developments; and to maintain the commitment and cooperation of federal agencies, private industries, state governments, universities, and most of all, the individuals who have to be involved to make it work.

Senator GORE. Thank you very much. We really appreciate that testimony.

Dr. Fields, about 20 years ago DARPA developed ARPANET, the first real computer network. And ever since then researchers have been finding new and unexpected uses for networks. Electronic mail which we take for granted today, as you noted, was invented on an ARPANET. Scientists are still busy finding new uses for networks today.

Today, we have networks with 1,000 times the capacity of the early ARPANET and S. 1067 calls for building a network 1 million times faster.

Could you describe some of the new applications that such a network might make possible? I know you demurred because it is, in fact, impossible to predict those applications, but can you—are there a few that you might envision being possible with a new high capacity network of this kind?

Dr. FIELDS. Well, all of the examples that we have worked out where we really can justify the high band width of the network because of the need for the application, and there are a lot of those, are the military examples which I started eluding to and I can go into in some greater detail if you would like it now.

If you really want more scientific civilian examples, I would rather that Bill answer the question. So it is your choice.

Senator GORE. Fine. Dr. Wulf?

Dr. WULF. The common thread that runs through almost all examples that people give have to do with visualization, the examples range from the sort of scientific visualization that we saw demonstrated today, through transmission of medical images to allow specialists at a remote site to help in the diagnosis of an illness. I am in some ways stumped just like Craig in that I suspect the most important uses of the high bandwidth are ones which we have not anticipated and will not until they happen.

Senator GORE. Last August, Dr. Wulf, you testified before this subcommittee about NSFNET and the need for more advanced networks. And since that time a number of things have happened. We have a new administration, a new budget and a lot of other new developments. Has there been much progress toward building a higher-capacity network since then?

Dr. WULF. There has been a great deal of progress along a number of different fronts. First, as chairman of the FCCSET subcommittee on networking, I requested the group of managers of networks in the various agencies, to create an implementation plan for the gigabit networks. They have done that. Those plans have been incorporated into an implementation plan for the entire FCCSET report which, as I said earlier, is complete and is in the approval process in the administration.

Just last week the National Science Board approved an expansion of NSFNET to include more nodes and more links to make it more capable and robust. We have further advanced on our plans to convert to 45 megabits during fiscal year 1990. So a great deal has happened.

Senator GORE. Well, very good. Dr. Fields, the OSTP report warned that Europe and Japan are aggressively moving ahead of the U.S. in a variety of networking areas with the support of concentrated government and industry research and implementation programs. Could you give us some examples of where the U.S. might be falling behind in networking?

Dr. Fields. Well, I do not have examples where the U.S. is falling behind in networking technology, but examples where the U.S. behind in networking implementation abound. Perhaps the most stunning examples are the current investments going on in Japan, and the purpose there is to promote high definition television.

The data we get from Japan is there investments are now running at about $600 million a year for promotion of high-definition television. And a large part of that is for infrastructure of high band width networks fiber optic networks to distribute signals.

Senator Gore. With what capacity?

Dr. Fields. I do not know how many gigabits, but it would have to be in that range. It is just a very, very large investment. There is nothing like that in the United States that I know of.

Senator Gore. I remember four years ago in another hearing on this topic, we had testimony about the Japanese science priority list and one of the top three priorities for that particular year was the development of a 10 gigabaud network. And presumably they have been moving forward on that intention.

Dr. Fields. I would just like to add, if I could just take a moment more, that while my comment about not being behind in networking technology is, I think, generally right, some of the critical components necessary from an implementation of a high band width networks up to electronic components—that is one technology area where we have fallen behind.

There have been any number of studies showing us falling behind the Far East in that area. And we might face a future in which we have to import the components to implement our high band width networks.

Senator Gore. One of the problems that we always hear about is the sad state of science education in this country. I personally believe that the national network can help to improve American science education. But I notice since you testified last August, Dr. Wulf, the National Research Network has become the National Research and Education Network. Why has this change been made? I applaud it, but why? And what does it signify?

Dr. Wulf. It is a recognition of what many of us had in our minds in the first place. I think that is the major reason. I think the potential for the network for use in education is absolutely enormous, not just in courses, although that is certainly true, but, in access to libraries and access to electronic laboratories, for example. We are beginning to see some of these uses now.

Again, the potential is just enormous.

Senator Gore. Do you call it NREN now?

Dr. Wulf. N-R-E-N.

Senator Gore. That is not quite the same, Doctor, I mean these are the really important questions.

Dr. Wulf. I have not heard anybody actually try and pronounce it; it is just N-R-E-N.

Senator Gore. That is all right. I will not press you on such a controversial matter. Dr. Fields, I think it is fair to say that we are all very eager to see this network built. And when I read the

FRICC report and when I then hear you say that the network will be operational in 1996, my first reaction is to ask, can we not build it more quickly than that? Why will it take seven years to get this network up and running? Can't we do it faster? After all, it is only going to take two years to increase NSFNET's capacity 30-fold.

If we continued at that rate of growth, we would have a three gigabit network in 1993.

Can we not accelerate this process? And should we look forward to an acceleration in the process?

Dr. FIELDS. Well, you can accelerate things somewhat by spending faster. You cannot accelerate them as much as you might like. By and large, I think that without any change in the current rate of investment, we are going to see these gigabit networks commonly available, at least I hope so, toward the end of the century.

You might be able to have that delay by an acceleration of investment like you are talking about. The issue is what else are you going to not going to do because of increasing that investment and that is a choice that is out of my hands. And that the Congress and the administration has to deal with that at other levels.

Senator GORE. Well, if we can cut it in half with the investment levels associated with S. 1067 I think that is clearly worthwhile because it is a very small investment. Could we get it even faster than that with more investment? Could we get it two years earlier with a reasonable increment to that investment?

Dr. FIELDS. It is conceivable but the uncertainties are just too high to say. Whether it is 1993 versus 1994, that is not something I think we can be precise about at this stage.

Senator GORE. Well we are going to give continuing attention to that question as these proceedings continue. There are a lot of other questions that I would like to ask both of you, but we have five more witnesses and we want to do justice to their testimony. They are very distinguished witnesses.

We really appreciate your attendance here today. Thank you very much for coming.

Let us go to our third and final panel now. Mr. Gray Collins, Senior Vice President for External Affairs at Bell Atlantic Corporation; Mr. Gene Gabbard, Chairman and Chief Executive Officer of Telecom USA, based in Atlanta; Mr. Richard Liebhaber, Executive Vice President of MCI Communications Corporation based in Washington; Dr. Robert Lucky, Executive Director for Research, Communications Sciences, at AT&T Bell Laboratories; and Mr. Roger Schwantes, Vice President for Production and Technology Development at Northern Telecom, Inc., based in Nashville, Tennessee.

Thank you all very much for coming. Without objection, your prepared statements will be included in full in the record. I am going to ask you to summarize your presentation in five minutes each. We do have time constraints at this point. And we apologize for the delay in getting to this panel, But I hope you have enjoyed the dialogue that has preceded you.

We are going to proceed with Dr. Robert Lucky as the first witness. And welcome, and please go right ahead.

STATEMENT OF ROBERT W. LUCKY, EXECUTIVE DIRECTOR, RESEARCH, COMMUNICATIONS SCIENCES, AT&T BELL LABORATORIES

Dr. LUCKY. Thank you, Senator. AT&T supports your bill and I——

Senator GORE. Thank you. We will go right on to the next witness.

Dr. LUCKY. I, personally, am very enthusiastic about it. I welcome your presence in this business. I think we need you. And frankly, I feel very frustrated. I am a member of a common carrier, but I also am a user. I am a member of the research community, and I have been really frustrated in recent years because I have a telephone at my desk and a separate computer at my desk. And on that telephone, I can call anybody in this room, anytime. I can speak as fast as I want in any language I want. If I do not know who you are, and I do not know a lot of you people, I can look you up in the telephone book. We are all connected this way.

You do not realize that in the data world none of those things apply. Not a single one of them. I cannot find any of you. There is no telephone book. You are not in it. I cannot talk as fast as I want. In terms of data speeds, I have a paltry little speed that I can use if I can even find you. I cannot talk any language I want because we have all these networks out there that all talk different languages.

And it is a tower of babble. So, frankly, I am really frustrated with this. And that is why I think we need your help in this business.

Now six years ago if we were doing this hearing, I would have felt personally embarrassed about this situation because as a member of the Bell System, I would have felt responsible for this. But in 1984, the Department of Justice relieved me of this embarrassment.

Senator GORE. We knew you would come to love it.

Dr. LUCKY. I can come here with a straight face and not be embarrassed because you cannot talk to your friends on the computer. Because we have a different paradigm now for running the networking business in our country. And it relies on competition between me and these other people at this table. And the problem is that paradigm is not creating the network and the innerconnection that we need.

So that is why I think we need the leadership of government. I wake up in the mornings with a nightmare that they are going to regulate this into existence, and that we do not want. But we need your leadership. We need you to stimulate this business.

Now, what AT&T has to offer particularly is the fiber optic technology. We are developing optical components and fiber optic systems. Today, they transmit 1.7 billion bits per second on a fiber. That is what we are installing today. That is about 50,000 voice calls on a fiber. Already in the research lab, we are sending 16 billion bits, 10 times as fast already in the research lab.

And so that optical transmission technology is doubling every year in the research labs. It is great technology.

But I want to caution you on that, Senator, because the idea is very prevalent that there is a lot of spare capacity out there in the network because of that fiber technology. Now my compatriots here are going to have to speak for their own companies. But I sincerely doubt that is true for AT&T.

The people who run the business have to compete at the bottom line and have to hold expenses down. They do not install a great deal of unused capacity at any given time. So it is like we have developed the technology, but it is like concrete ready to pour. It is not poured out there. And it takes the stimulation of something like your bill to get us to pour that capacity.

And then after you have poured the capacity, you have the roads. Then you have to build interchanges. Now, again, in the research lab, we are working on optical switching and very high-speed switches that could switch these tremendous bit streams that are out there. So the next thing you need is the interchanges; then you need the on and off ramps.

And you need a motor vehicle code. You need someone to administer this and to pull it all together and say that you are not going to speak different languages. You need leadership force, the administrative structure, and rules of the game. Someone has to pull it all together, and competition is not doing it.

Finally having welcomed you to this and saying why we need you, let me say that it is a chicken and egg problem. You put it very well. The technology is there, but the market is not. And that is why we are not doing it. You might say that if this is such a good thing, why isn't AT&T doing it? Because the business people do not think there is money in it right now. But there will be, and there can be a tremendous market if you get out there and stimulate it for us. And then you get out of it. Okay? [Laughter.]

[The statement follows:]

Statement of

ROBERT W. LUCKY

Executive Director - Research

Communications Sciences Division

AT&T BELL LABORATORIES

AT&T supports government investment in the technology and availability of national data network services, as proposed in the High Performance Computer Technology Act of 1989. Availability of such services would promote scientific research and collaboration, and help make U.S. technology more competitive.

In general, computer networks today are not interconnected and it is often impossible to send information between any two systems. Also, computers require much faster transmission speeds than voice communication. Many businesses, most campuses and research organizations have local area networks that transmit at speeds of about ten million bits per second, and new networks are being developed that transmit at 100 million bits per second. There are also generic uses for data rates much higher than those available in today's wide area communications networks; transmission of pictures or graphics are examples of such uses. But the most important uses of a high-speed network are very likely to be uses we have not yet identified or even imagined. We should not be so shortsighted as to try to justify the nation's investment in such a network solely on the basis of today's uses.

Inter-city optical transport facilities are currently receiving adequate attention, but other important parts of the network are not. What needs further development is an overall architectural plan, which incorporates multiplexing arrangements that enable data users to share the transmission capacity effectively, and a logical structure (directories, network management plan, etc.), which embodies all the rules for the use of the network.

To stimulate the investment needed for a successful high-speed data network, the buying power of data-networking services should be put in the hands of the users, as it is for telephony and travel. We agree that networking for supercomputer users is strategically important. At the same time, we believe that our country should conduct research that can lead to data services that meet the broader needs of industry and government.

The proposed gigabit network would, in our view, stimulate scientific discovery and benefit the U.S. economy and the American people in many ways. Optical-transmission technology does not require additional government funding. The research sponsored by this bill should be directed toward the infrastructural requirements, particularly overall architectural design and network management.

My name is Robert W. Lucky. I am Executive Director - Research in the Communications Sciences Division of AT&T Bell Laboratories, at Holmdel, New Jersey. I wish to thank you, Mr. Chairman, and the Committee on Commerce, Science and Transportation for the opportunity to present AT&T's view on high-speed computer networks and on your proposed legislation.

Government investment in the technology and availability of national data network services, as proposed in the High Performance Computer Technology Act of 1989, would promote scientific research and collaboration, and help make U.S. technology more competitive. As the use of the resulting network technologies spread, they would also enhance commerce and promote efficiency in business operations. In order to achieve these high-level objectives, the network will need to exhibit high user speeds, be broadly available and friendly in access. All this must be achieved at affordable user costs and be profitable for service providers.

Today's telephone network serves our needs for voice communications quite well. We can call anyone, anywhere. If we do not know the person's number, we can look it up in the telephone book or call directory assistance. Once a call is connected, we can talk as fast as we are able and we can use any language we wish. Since we take these attributes for granted, it is hard to recognize that almost none of them apply to data communications networks that link together computers and terminals.

In general, the computer networks of today are islands unto themselves. They are not interconnected and it is often impossible to send information between any two systems. There is no directory of computer users and no one appears to have either the responsibility or the desire to provide one. Furthermore, computers need to be able to talk much faster than the voice telephone network, upon which they rely, now permits.

Why are higher speeds necessary? A century ago we had a nationwide digital network for telegraphy. Users couched their messages in "telegraphese", conserving words wherever possible. Anyone wanting to send a precious message had to trudge down to the telegraph office where the message was keyed into the network by expert operators at rates of about ten bits per second. People in that era thought that was all the communication speed that would ever be needed. It was a speed comparable to human speech -- what more could anyone want?

In the 1960's, modems were developed to convert computer data into an analog, speech-like signal for transmission over the telephone network. Today, modem speeds of 2400 bits per second are typical. At this speed, a typical printed page is transmitted in about ten seconds. While that is much faster than the average reading speed, it is somewhat slower than the rate at which we skim material, looking for specific information.

Although 2400 bits per second is suitable for human reading and typing, it is inadequate for most machine functions. A computer's efficiency depends on its ability to move large amounts of data from place to place. Even personal computers today transfer data to and from their internal memories at effective speeds of about a half billion bits per second.

As a simple example, consider today's popular word-processor programs. In recent years even these relatively simple programs have ballooned in size as more and more functionality has been added. Today, a typical program might be one megabyte (eight million bits) in size. If this program is stored external to the computer — as is the case in more and more business applications where workstation computers are served centrally by shared-storage systems — then a user will be affected by the length of time it takes to retrieve this eight-million-bit file. At 2400 bits per second, the user would have to wait approximately one hour for the program to be loaded.

In short, for a useful exchange of programs and data files between computers, higher-speed networks are essential. Thus many businesses, most campuses and research organizations are served by local-area networks that transmit at speeds of about 10 million bits per second. (The actual transfer rate of data is considerably less than that, because of the bottleneck in communications processing — a subject requiring further research.)

New networks are being developed and standards are being readied for local-area networks that transmit at 100 million bits per second.

But when the computers on these local networks exchange data with other computers outside their own locality, they will be reduced to paltry rates of several thousands or tens of thousands of bits per second. A fire-hose stream of bits in a local-area network will become a drinking-straw stream when it goes beyond the local network.

You and I too can make use of higher bit rates. There is an old saying that a picture is worth a thousand words. Curiously, it requires almost exactly a thousand times the communications capacity to transmit pictures as spoken words. Network television signals, for example, are transmitted digitally in the telephone network at 45 million bits per second as compared with the standard 64 thousand bit-per-second rate used for digital speech.

A still picture, when digitized, is transformed into a file of about 24 million bits. But people have a habit of leafing through pictures at a rapid rate. As we turn the pages of National Geographic, for example, we may be using the equivalent of hundreds of million of bits per second. The traffic in medical imagery within hospitals, as another example, reaches comparable figures. Every waking second our own optic nerves carry several gigabits into our brain.

Thus it can be seen that there are generic uses for data rates much higher than those available in today's wide-area communications networks. Generally speaking, we need these higher rates not because we have a continuous demand for huge amounts of data, but more typically because we need to send data instantly. We may only need the channel sporadically, but when we have information to transmit, we require the full speed of the network for our individual use -- we cannot afford to wait.

Finally, in my view, the most important uses of a high-speed network are very likely to be uses that we have not yet identified or perhaps even imagined. In all past telecommunications history, users have been satisfied with the existing speed of transmission until they were given higher speeds -- then and only then have new opportunities developed. Only in retrospect have the previous transmission rates looked inadequate. Investing in our national data network infrastructure is an investment in our future. It is like building a road into the wilderness. We must not be so shortsighted as to try to justify such an investment solely for today's uses; these uses grew around a lesser system.

Given that a high-speed national networking infrastructure is needed, what elements are already being developed and what elements need to be stimulated by the government? AT&T is investing in a great deal of research and development directed towards achieving a very high-speed optical-fiber transport system. Progress in optical transmission technology has been dramatic. The fibers that we are installing in our network today carry data rates of 1.7 gigabits per second in each fiber strand. Already we have experimental systems in our research laboratory that transmit data at 16 billion bits per second in an optical fiber. For the future we expect continuous growth in this capacity. We believe that the ultimate capacity of the fibers already installed in the nation's backbone may be as high as ten terabits (trillions of bits) per second in each fiber.

Of the elements needed to implement a nationwide, very high-speed information network, only these inter-city optical-transport pipes are currently receiving adequate attention. It is as if we had built superhighways between our cities, but no interchanges, or on-and-off ramps for access. What needs further development in such a network is an overall architectural plan that incorporates the multiplexing arrangements that enable data users to share this high-speed pipe effectively, and the logical structure, e.g., directories, network management plan, etc., that embodies all the rules for use of the network. To further extend the analogy to the highway system, in addition to the access ramps and interchanges, we need a motor vehicle code that tells everybody exactly what kinds of vehicles are allowed and how the roads are to be shared.

Investment in the infrastructure needed to apply high-capacity fiber-transmission technology to specific data applications, e.g., supercomputer users, is driven by market demand and prospective return on investment. This, in turn, is driven by our ability to meet the needs of users economically.

Two barriers must be overcome to create the conditions that will stimulate increased investment in the infrastructure needed to serve researchers in general, and supercomputer users, in particular. These are: (1) data network services need to be valued like other researcher support services, and (2) shared services must serve a variety of users with different bandwidth demands, if it is to be marketable and profitable.

In other words, data communications services must be treated as one of many services that researchers purchase from their grant funds. The research community would not think of creating its own airline or phone system or postal service. Now that data communications services have become an integral part of a researcher's requirements, it is important that the purchase of these products and services be made from the marketplace and that the purchasing process be subject to the same forces of the marketplace as voice services and travel services.

One way to accomplish this is to change the flow of grant funds for networking. Instead of channeling funds directly to the network providers, e.g., the regional networks, these funds should be incorporated as part of research proposals or included as indirect costs of researchers (or universities acting on their behalf). This is the model for voice services and should serve as a model for creating competitive and profitable price levels for data services, as the Federal Research Interagency Coordinating Committee (FRICC) report has recognized.

If the purchasing model for data services is changed to create price levels that are based on market value, as suggested above, the remaining question is how to serve the demands of supercomputer users at reasonable price levels. Because supercomputer users' needs may at critical times exceed by more than a thousand times the needs of other users, the question of how to price usage affordably needs to be answered.

The peak users, e.g., the supercomputer users, will drive the instantaneous peak demand of the network. This, in turn, will determine the instantaneous peak capacity required of the network. If the supercomputer user is the only user, then the total cost of the network will be carried by this one class of user. If, on the other hand, a large number of users are able to access and use the high-capacity data highway, the cost of the network will be shared by a large number of users. For this to happen, supercomputer users (and others) must be able to take control of a required amount of the bandwidth at a premium price.

The price for services on a more broadly shared network would be substantially less than the price for services on a network that served only supercomputer users. This is because the revenue of the lower-priority users would come without additional investment in the highway—albeit by increased investment in the infrastructure to allow sharing and priority treatment of many users. Research is being performed at AT&T Bell Laboratories and elsewhere into techniques for sharing high-speed networks in this manner.

In our view, the ability to commercialize a gigabit network will be enhanced by creating a mass of lower-priority users who would purchase data services from the same highway but at priorities and prices that serve their needs. One source of this traffic is the operational and administrative traffic currently being handled on the internet, e.g., electronic mail.

Other possible sources of traffic to "fill in the valleys" of demand when supercomputer applications are not active are the data communications requirements of other government users. If the infrastructure built to support the gigabit highway were created in such a way that it could serve the broader needs of the government, the additional lower-speed traffic would be able to support a network service with lower price levels for supercomputer users.

Since the growth of supercomputer traffic will be based on scientific discovery, it will be several years before supercomputer users require a substantial part of the total available capacity of a multiple-gigabit highway. It is necessary, therefore, that some substantial traffic source be able to take advantage of the highway in the near term. Broad government requirements could help serve this need.

Our conclusion is that the proposed gigabit network would stimulate scientific discovery and benefit the U.S. economy and the American people in many ways. Further, the optical-transmission technology needed to create the transmission facilities for gigabit transmission does not require additional government funding. The research sponsored by this bill should examine the infrastructural techniques needed to allow traffic from multiple users to be served by the gigabit highway. In particular, research is needed into an overall architectural design that meets a variety of price/performance options for end users and addresses the broad requirements of network management on an end-to-end basis. Such techniques should be studied and targeted for inclusion in future plans to serve the government at large, and the research community in particular.

Senator GORE. That is a deal. I am glad I did not cut you off prematurely.

We will just go right down the line here. Mr. Richard Liebhaber, Executive Vice President of MCI, is next. Welcome.

STATEMENT OF RICHARD T. LIEBHABER, EXECUTIVE VICE PRESIDENT, MCI COMMUNICATIONS CORP.

Mr. LIEBHABER. Thank you, Senator, for the opportunity. On behalf of my colleagues at MCI, I would like to start with two adages.

The first one is that in real estate, location is the key. In networking, infrastructure is the key. The second adage is a good idea at the wrong time is a bad idea. This is a good idea. And we believe the time is right at MCI.

Let me try and put the role of competition and technology into perspective for you from our view, and deal a little bit with the economics of networking that we have all experienced in the United States.

1972 a voice grade private line, if you were to acquire one, would cost you $1 per mile. At best it would have carried 1,200 bits per second. Unreliably at that.

Senator GORE. What was that?

Mr. LIEBHABER. About 1,200 bits per second. But a 1972 dollar today is about three current dollars in 1989. Today you can purchase a T1 or a 1.5 megabit private line for less than $3 per mile. That is 1,250 times price performance improvement in 17 years.

In 1984 it cost interexchange carriers, such as MCI approximately $1,000 per circuit mile to construct 64 kilobit circuit mile. In 1989, our cost at MCI are under $1, or three orders of magnitude improved. Incredible, incredible technology and price performance benefits.

We see the conversion of electronics to 1550 nanometer from 1330 as a result of our colleagues at AT&T and Northern Telecom. We see the development of coherent lasers and new silocondopent materials. We see further improvement in band width and repeater spacing. All leading to better cost performance, better price performance.

At today's levels, 45 megabit, the capability of which we are exploring for the National Science Foundation Network Phase II, we see that transmission cost at 45 megabit being approximately $1 a second. To go back to your example regarding encyclopedia's, we could transmit one volume of the Americana encyclopedia in one second, approximately 50 megabits for $1, irrespective of distance.

Assuming conservatively a 20 to 30 percent compound growth rate improvement, which is what we have been experiencing in overall price performance in three years the cost of 50 cents per second at 45 megabits or lower is virtually in the cards.

From that perspective, this is a good time. Now is a good time to pursue broad band networking.

I have been asked to focus on the benefits, and I come to you as sort of an artifact in this industry. Without revealing my age, I would like to just brief you on the fact that supercomputers are a long history in my blood. I go back to the 7030 stretch in the late 1950s produced by a manufacturer I am sure you are familiar with.

And after spending a number of decades in that industry, for the last six years, I have had the opportunity to deal with the network side of this melding of computers and communications, something which we at MCI call the information age. We have over 16,000 route miles of high performance fiber performing at 1.7 gigabits in some areas and dealing with cross sections today of eight to 10 gigabits in our network.

We are interested in supporting the development of research and education networks by aggressively competing and providing people like NSF not just with facilities, but with intellectual assets as well. As Dr. Wulf described, the demand growth on that network is phenomenal, 30 to 35 percent compound growth rate, which is causing us to expand that network at a very fast rate.

As we understand the bill, it would authorize the construction and operation of a three gigabit per second switch network to link 1,300 institutions representing government, industry and higher education. And, Senator, we support that enthusiastically.

From our perspective we see no regulatory barriers that would prevent the execution of this project. We see no regulatory barriers either on the long-haul portion of the network or in the interconnection of the network at the 1,300 institutions.

But networks are made up of more than just optical fiber backbones, as you know, they include switches and management systems and access facilities and numbering plans and billing systems and traffic systems. To gain experience in these and other aspects of the provision of network services for large computer applications we did bid with Merit on the National Science Foundation network. We have won that bid and are participating in that growth.

We encourage you to explore further application, further widening of that band width, and further spreading of the notion of the information society.

[The statement follows:]

STATEMENT

OF

RICHARD T. LIEBHABER

EXECUTIVE VICE PRESIDENT

MCI COMMUNICATIONS CORPORATION

Mr. Chairman,

Thank you for the opportunity to address your committee on the National High-Performance Computer Technology Act (S. 1067) which you recently introduced. The work you are undertaking is important to our country, our industry, and to MCI.

I have been asked to focus on (1) the benefits that a National Research and Education Network will provide to the telecommunications industry and other industries and (2) how best to design and build such a network.

Let me first explain the perspectives I bring to this issue. They are several. Most of my career has been spent in the computer industry with IBM, in which I was involved in the engineering of very large distributed processing computer networks and in the applications that utilized those networks. For the past several years I have been at MCI in charge of the Engineering and Operations of our multi-billion dollar network and the information systems needed to support our business. MCI is a company which, to a large extent, created the competitive telecommunications business. It was also the company that pioneered the implementation of single mode optical fibers in national networks. I believe we still maintain the leadership role in advancing the use of higher and higher speeds in our network. We have over 16,000 route miles of single mode fiber, over 400,000 fiber miles, some of which are operating at 1.7 gigabits per second. This is not experimental. This is carrying high speed computer data, images, and voice traffic. Furthermore, we have shown our interest in supporting the development of research and education networks by aggressively, competitively providing the National Science Foundation with an improvement of over 20 times in the performance of their network.

To summarize, we have a very real appreciation for the users and engineering of high performance computer networks and a strong belief in the value of competition in telecommunications. We have made major investments in advancing digital communications technology and in providing the advantages of those advancements to the general user through significantly lower rates and to the National Science Foundation to improve the performance of its network.

My remarks in this short time will be restricted to Title II of S. 1067 regarding the National Research and Education Network.

As we understand the bill, it would authorize the construction and operation of a three gigabit per second switched network to link 1300 institutions representing government, industry, and higher education. We support this initiative.

We, the interexchange carriers, have over the past four years spent billions of dollars to construct fiber optic digital networks which provide data rates far exceeding the requirements outlined in the proposed bill. Just one of our typical fiber bundles of 11 pair operating at 1.7 gigabits per second will today provide 10 times the capacity your bill desires. And this is just MCI. At least two or three of our competitors have similar cross section networks. This capacity has been constructed at no cost to the government or the taxpayers. In fact, it has resulted in significantly reduced long distance telephone rates. A reduction of over 40% since divestiture! This committee would probably not be considering this legislation if the common carriers had not made these large investments. Today we can provide the bandwidth, as can our competitors. The government can procure the transmission capacity it needs under competitive conditions, yielding the best allocation of all our resources. This is not a chicken or egg problem. We have already built the chicken.

But networks are made up of more than optical fiber backbones. They include switches, management systems, access facilities, numbering plans, billing systems, traffic systems, etc. In order to gain experience in these other aspects of the provision of network services for large computer applications, MCI bid as a subcontractor to MERIT, for the NSFNET. We have won that bid and are participating in the growth of the network.

However, NSFNET is not the only activity in which we are involved to develop new services for computer networks. We have work going on in our laboratories and in the laboratories of our vendors to develop very fast, very high speed nodal equipment for non-voice information transfer. I expect our competitors have similar programs. We would like the opportunity to offer these new competitive services to the proposed National Research and Education Network, and to explore their application and further development.

Through our work with NSFNET, as well as research work with commercial and government customers, and with our colleagues worldwide, we believe we have gained an understanding which will yield innovative products to serve much of the expected needs for data, voice, and image communications, including supercomputers.

We believe these varied needs can best be served by the energetic interplay provided by an open, free, competitive system.

To use an analogy with which the Chairman is familiar, the interstate highways for supercomputer networks have already been built -- at no cost to the taxpayers or the government. The free enterprise system has done its job. The legislation should focus on the application and use of that infrastructure. Through the consolidation of the research and education communities, we envision an environment in which we can concentrate the development process on one wideband network. Like our work with the NSFNET, we are prepared to share the intellectual assets that result from the efforts. The best way is for the users of supercomputers to identify the freight they want carried -- develop the application set.

In some cases the supercomputers will have large loads which can be batched and transported very quickly, equivalent to an air freight shipment. Other applications will be continuous which will be more analogous to a gas pipeline. Still others will be smaller and more distributed cargoes, most appropriately served by the equivalent of trucking companies. Some will be local like the taxicab market. The critical challenge to be addressed here is the infrastructure, control and management disciplines required. Dilution of sub-network or agency specific networks will not in my opinion assure such development. Under the leadership of NSF we envision such a plan bringing the broad application requirements together, thereby assuring critical mass traffic.

The major problem today is that the users have not quite figured out how best to pack their freight and we carriers have not quite figured out how best to manage and switch the cargo. Modest efforts like the present NSFNET are appropriate ways of working together to gain some mutual experience.

All of the major carriers are contributing to the worldwide attempt to reach standardized solutions to these traffic management problems. We are not yet of one mind on finding the best solution. Some are looking for the universal, do-everything switch; others are focusing on individual application areas within the supercomputer universe for specialized solutions. Some of these developing services will match markets, some will not. We believe the market is the place to test these ideas. But the stimulation contemplated by this bill could provide the incubator for such a process. The issues facing the designers of supercomputer networks are efficient allocation of resource issues and management issues. We already have the first component in place, the greatest multi-billion dollar laboratory in the world: the competitive telecommunications industry in the US -- a model the rest of the world is now trying to emulate.

Referring to the draft Program Plan for the National Research and Education Network, we suggest that the plan be allowed to progress to Stage 2, the shared 45 megabit per second trunk system, as soon as possible, and that this committee encourage joint development and cooperation with the NSF by the appropriate user agencies, NOAA, DOD, NASA, etc. We all need to take risks.

Beyond Stage 2, the trade-offs to be made will be the classic communication decisions of increased bandwidth versus terminal processing.

To summarize, our experience in computers, telecommunications, NSFNET, and competition leads us to conclude that the rapid development of the supercomputer industry will be furthered by this bill through the free market bidding process in response to the consolidated NSF process we suggest.

Thank you for asking us to share our views with you on this important issue. I would be pleased to answer any questions the Subcommittee may have.

Senator GORE. Thank you very much.

Mr. Gray Collins, Senior Vice President for External Affairs at Bell Atlantic will be our next witness. Senator Pressler is going to run the hearing for a moment and I will be back shortly. Please proceed with your statement.

STATEMENT OF A. GRAY COLLINS, JR., SENIOR VICE PRESIDENT, EXTERNAL AFFAIRS, BELL ATLANTIC CORP.

Mr. COLLINS. Thank you, Senator. I would like to thank the subcommittee for inviting me to appear, this afternoon and provide Bell Atlantic's views on development of the National Research and Education Network. Bell Atlantic is critically interested in this area.

We built local fiber optic highways to link colleges and universities in our region with supercomputers, and we have recently filed a waiver with the decree court asking, in part, to provide supercomputer services on a timeshare basis over our network.

We know how challenging and difficult it will be to build the network. The technology for this super high-speed data highway does not even exist today.

One approach is for government to provide funding for research and development. Funding can play an important role, but the best way to get the job done, at least in my opinion, is to turn to the private sector. The government has established 1996 as a day in which the network will be built, but it would be—it should really solicit bids from the private industry to build that network. And ownership and operation of the network would remain in the private sector.

To get the best network as quickly as possible and at the lowest cost, that bidding should be as competitive as possible, Senator. That means that there should be no artificial restrictions on the company to—companies who want to bid for example the Bell companies with their vast expertise and sophistication in telecommunications should be able to bid on this long distance network without regard to the long distance restrictions imposed on it by the AT&T consent decree.

There is a precedent for this kind of relief. When the government sought bids on the new FTS-2000 network, the decree court allowed the Bell companies to bid to provide service that according to some were long distance services. And that grant to allow us to participate, ensured that there was real competition in the FTS-2000 bidding. The FTS-2000 network is a complex one. But it pales by comparison to the super highway for supercomputers proposed in this legislation.

That is all the more reason that the bidding for the super highway should be open and as competitive as possible.

In addition, the competitors need to be free to experiment and develop the technology that will make the network possible. That means the decree's restriction on manufacturing, which inhibits research and development, must be eliminated, at least for this project S. 1067 unfortunately is silent on the AT&T consent decree. I would suggest that Congress must address these issues if you are to achieve the goals articulated in your proposal.

I would like to comment briefly on what the network should look like. It should not be a network that links up just a few locations and can be accessed by only a few government locations and big universities. It should be part of the public switch network, thereby accessible to the wide variety of users including small and medium-sized businesses and small colleges.

Such an interconnectable network would obviously make supercomputing resources available to the widest possible audiences, including the rural areas. And it is an important public policy goal, but it would do more than that. The existence of this high speed fiber optic network would provide the incentive for the local telephone companies to install fiber in their own network so that the small and medium-size business can be reached over the super highway.

Today, the telephone companies are installing fiber in their backbone of their local network. But this project will give them the incentive to expand the fiber to the customer's home and businesses, the so-called last mile and bring the benefits of the information age to the American consumer.

Congress can encourage this important local deployment of fiber by other means as well. For example, local fiber systems are the cutting edge technology for the distribution of cable television services. By lifting the ban on telephone company provision of cable television, Congress will give added incentive to the telephone companies to bring fiber to that last mile.

Providing these incentives is important because bringing fiber to the last mile is not going to be easy. The technology is not yet available. We have to develop higher speed commercial grade optoelectronics and larger and higher-speed switching suitable for remote terminals and end office environments. With the right incentives, however, we are confident that the job will get done.

It is an exciting time. The National Research and Education Network will be a powerful vehicle for making supercomputing services more widely available and for encouraging the development of the commercial high-speed networks envisioned by this legislation. With the help of Congress, the telecommunications industry including local telephone companies, can make that promise a reality.

Thank you.

Senator PRESSLER [presiding]. Thank you very much. Later I would like to have you comment on how realistic it is to access small businesses and what their benefits will be.

But first we will hear from Mr. Gabbard.

STATEMENT OF O. GENE GABBARD, CHAIRMAN AND CHIEF EXECUTIVE OFFICER, TELECOM*USA

Mr. GABBARD. Thank you, Senator. I appreciate being invited here today. Telecom*USA is the fourth largest long distance telephone carrier in the United States, of course, after the breakup of AT&T.

Our company owns and operates a 3,000-mile digital network, primarily fiber-based, which is part of a 19,000-mile nationwide system operated as the National Telecommunications Network.

And NTN is a partnership of six regional companies in the United States.

And I certainly agree with Dr. Wulf, that the issue is really not whether information is going to be important or not or whether we need this network or not, because to effectively move, manipulate and use information is literally the key to the prosperity of the United States, certainly in my opinion, Senator.

We need a clear, coordinated national policy for high-speed networking of supercomputing. And I commend you, Mr. Chairman, and Congress, for recognizing this challenge, and believe Senate bill 1067 lays the foundation for such a national policy.

And I might digress here for one moment to say that I think that one of the greatest strengths that we have in the United States is an infrastructure for funding and starting new ventures, an entrepreneurial infrastructure. And that very infrastructure, which creates lots of new ventures that are very aggressive, just like our company has been in putting in fiber in this country, is an impediment in terms of organizing a big national system like this. And that is one reason why we need government support to help guide and to help set a national policy.

I think the development of the National Research and Education Network, and I will call it NREN as well, Senator Gore, is really an exciting project. And it is something that the various entities, many of them, can use to pull around, to focus on, to really get underway and benefit all of America.

On the implementation of the network, I would like to focus my comments really on three areas. The first is, and I am agreeing with Dr. Wulf again, that we need to move quickly and effectively, and it is very, very critical that government, the telecommunications industry, computer and data networking industries, be tightly linked in a common force, operating toward the same goals, standards and directions. Otherwise, I believe at least two standards are going to end up again; you know, one that is tailored toward institutions and government directions, and one tailored toward industry. And I am afraid it might take a long time to bring those together and make them work in an optimum way.

I also believe that new structures, when you have a large new task at hand, can be awfully efficient. Because they are not burdened with all of the red tape that existing structures have. For this reason, Mr. Chairman, I really propose that a new not-for-profit corporation be established with a strong executive officer and a board of directors; and that that board be made up of users in government, half of it, and that the other half of it be made up of the industry segments that have the expertise needed to move this project out quickly. Supercomputing, long-distance telecommunications, local telephone companies, that expertise is needed from all of those points.

Senator **Gore** [presiding]. So you would like them to run it instead of NSF; is that what you mean?

Mr. **Gabbard**. Yes, sir.

Senator **Gore**. Well, let me say that we are open to that. I do not want to interrupt your testimony prematurely, but we are open to that idea. And it can be done even after the legislation is passed, but we will actively consider that at this stage of the process.

Frankly, who is put in charge was a question that we thought about at some length; and there are different opinions. And we are going to have continuing discussions on this as the bill works its way toward mark-up. But I am happy to hear your opinion on it.

Excuse me for interrupting.

Mr. GABBARD. Thank you, Senator.

I would add a corollary to that, that virtually all of the work should be done in existing government entities and industry. And I think that DARPA is doing an outstanding job, for example, of developing new packet switch technology, and that should continue. But I am worried that if you manage it from within one of the present structures, that it will go slow because of all of the history that resides within one of the present structures.

Senator GORE. Of course, they contract it out, and might take that approach in the larger network. But, excuse me again. Keep going, and then we will come back to this.

Mr. GABBARD. Okay.

I think equally important, or perhaps more so, that the new corporation that I propose should be required to plan its own demise. And, in fact, I would even go further and say that the charter of the new corporation should require this network to be fully operational at the 3 gigabit level by the end of five years, and be transferred back to commercial operation within two years thereafter, so that we have definitive goals.

If we could put a man on the moon in one decade, Senator, we should be able to build this network, which is a much smaller project and where we have lots of technology, in half a decade.

My second recommendation has to do with technological implementation. And I certainly agree with Mr. Liebhaber here, there are three networks and a fourth underway, certainly, if you include the AT&T network, there are three seamless fiber networks that cover the entire United States in the 16,000-and-higher-mile range each, they go to lots of different places, so it is not just duplicates of each other. So the transport mechanism is here now. We do not need to worry a heck of a lot about that, if at all.

The packet switching that we need, new, fast packet switching, is being worked on in lots of places, and I believe 5 gigabit machines, based upon what I am being told, will be available within the next three-year time frame from industry, and also though, thanks to DARPA and their support in that area.

What is missing, and I believe the key area technically, are the protocols, the new protocols for addressing routing, converting, to provide easy access, very high reliability through the network on alternate paths, self-healing; and that is where a lot of the focus should be, in terms of the money and the effort from a technological standpoint.

We have great protocols for lower speed, but we do not have protocols that are friendly and easy and work at the higher-speed areas.

Third, Mr. Chairman, I think we ought to get on with it. I think everything is in place. Americans have proved time and again that they can get the job done if given the incentive, and if we create a streamlined organization to go. And in the meantime, I think we ought to go ahead and use the existing fiber networks to hook up supercomputers today with multiple 45-gigabit links so that we can see what happens today, because the ability to do that is here today. And that will also help in the development of the protocols and proving of the protocols for the larger network.

So, in summary, Mr. Chairman, I believe that a 3-gigabit network can be fully realized in five years. I applaud you, the other members of this subcommittee, and Congress for pushing us in the right direction down the path.

And I thank you for the opportunity to appear here today.

[The statement follows:]

COMMENTS OF
O. Gene Gabbard
Chairman and Chief Executive Officer
Telecom*USA

Before the
Senate Subcommittee on Science, Technology, and Space

June 21, 1989

Mr. Chairman, Members of the Committee;

My name is Gene Gabbard, Chairman of Telecom*USA, the fourth largest long distance telecommunications carrier in the United States. The company operates a 3,000 mile digital, predominately fiber optic, network, and provides long distance service in thirty-two states in the Southeast, Midwest, and West. The company also sells telephone systems, provides data base management services, operates an independent telephone company, and publishes telephone directories.

1. Introduction

We are in the midst of an exciting era. Our society is changing from the manufacture and exchange of goods to one that relies on the creation and dissemination of INFORMATION. The key to our national prosperity rests in our ability to effectively move, manipulate, and use this information.

The United States must have a clear, coordinated national policy for the harnessing of the information age. Other nations either have or are planning such a policy. During the past five years, countries that were served by outdated and inadequate national telecommunications systems have made tremendous strides. The telephone system in France was once the subject of jokes. Today, France is recognized as a leader in telecommunications systems. Japan had been thought of in the authority to get the job done. We must have a strong leader with the courage and ability to implement the NREN in an excellent manner in a short period of time.

2. Organization

Providers, users, and developers of the multi-gigabit NREN must not be isolated. They should be grouped into a single well-coordinated force. It is also crucial that the involved Federal Government agencies, the telecommunication industry, and the computing and data networking industries be tightly linked into a common force with the same goals, standards, and directions. Otherwise, at least two standards will be developed, one tailored along federal and institution lines, and one along commercial lines. It is clear that the NREN is too important to allow its management to be dominated by a single group of players.

For these reasons, Mr. Chairman, I recommend that a separate, not-for-profit corporation be established with a strong executive officer and a board of directors. The makeup of the board of directors should be one half from the Federal Government and user groups, and the other half from the industry groups needed to supply technologies and services. The Board of Directors should have final authority within the law on the approval of standards, selection of technologies, selection of suppliers, disbursement of funds, appointment of officers, etc. just as in any corporation.

The new corporation should be kept lean, with virtually all of the work, except overall management and control, being past as a follower in advanced information processing and movement. It is from Japan and France that many of the American fiber optic based telecommunications carriers obtain portions of their transmission equipment. Innovation in obtaining gigabit per second speeds over a single fiber pair has been more readily and economically available from the suppliers of these nations than from North American manufacturers. For the United States to maintain its role as world leader in information technologies, scientific advancement, scholarly research, and the production of excellent scientists, engineers and technicians, we must have a clearly articulated national policy covering the movement of information. Telecommunications, the products of scientific advancement, and a well trained, skilled populace are national resources of the next century. The National Research and Education Network (NREN) is a key element of a national telecommunications policy.

We can no longer separate telecommunications, computers, software, and research into isolated categories. They are irreversibly connected. So, too, are the providers, users, and developers of new technologies. Our efforts must be directed towards creation of continuity and orderly progress. As detailed in S. 1067, the President, through the Federal Coordinating Council on Science, Engineering, and Technology (FCCSET), is responsible for successfully addressing the needs of national high performance computing research and development efforts. FCCSET, or whomever ultimately is charged with the task, must be given the tools to manage this process tightly, and held accountable for its results. They must have the carried out by existing government entities and industry. For example, the preparation of standards should be done jointly by the National Institute of Standards and Technology and the Institute of Electrical and Electronic Engineers, under the total control of the new corporation. The completion of the development of new fast packet switching technology should be managed by the Defense Advanced Research Projects Agency (DARPA) through the competitive bidding process with at least two or three awards to industry. Such contract awards should be on firm fixed price and firm schedule terms, with industry sharing the risk. Multi-gigabit transmission facilities should be obtained through the competitive bidding process from existing fiber optic network operators.

Perhaps most importantly, Mr. Chairman, the new corporation should be required to plan its own demise. I believe the NREN should be completed within five (5) years, to at least the degree permitting handoff to commercial carriers. I, therefore, propose that the charter for the corporation call for full operation for five years, followed by a two year wind-down. The major work during the wind-down would be the smooth handoff of network control and operation to the commercial carriers, to not one but at least three. Otherwise we could find monopoly or duopoly control over prices and services after handoff to industry.

One will recall the great success of the National Aeronautics and Space Administration (NASA) when it was established as a fresh new entity with a clear and concise mandate. Although much smaller in scale, the NREN project will provide focus to American business, research and education not seen in recent times. The spin-off and positive side effects will be substantial, in addition to ensuring the completion of a national resource in a short period of time.

3. Network Implementation

The implementation of regional and specialized networks such as the Southeastern Universities Research Association Network (SURANet), the Energy Science Network (ESNET) sponsored by the Department of Energy (DOE), the NASA Science Internet (NSI), the New York State Educational Research Network (NYSERNET), etc. have been highly successful and are significant national resources. These networks, through the efforts of the NSF, at the direction of the Executive Office of Science Technology Policy (FCCSET), and more recently, the Federal Research Interagency Coordination Council, (FRICC), are being linked into a national network operating at "T1" (1.544 Megabits-per-second).

Work is underway to upgrade to "T3" (45 Megabits-per-second) connectivity between gateways, as part of the National Network Testbed (NNT) and the Research Interagency Backbone (RIB) projects, sharing facilities among DARPA, DOE, NASA, NSF, HHS, NOAA, and other federal agencies. Telecom*USA and our joint venture partners in the National Telecommunications Network (NTN) have actively participated with and provided services to both SURANet and NYSERNET. Using the 18,000 mile NTN network, we are already providing the bandwidth necessary for the NNT.

These projects should continue at full speed. However, the multi-gigabit NREN should be the final carrying network for all of these regional and specialized sub-networks. Therefore, close coordination must be established on day one among the planners, builders, and operators of these networks and the new corporation. The organization proposed for the new NREN corporation should facilitate this transition.

The new multi-gigabit NREN must be a "new design" because of today's limitations in high level digital multiplexing, packet switching, and networking interfacing, routing and control protocols - both hardware and software. However, the new network must be "upwards" compatible so that all of today's terminals and computers can operate over and in harmony with the new multi-gigabit NREN.

I am not an expert in data networking. There are many devoted experts within many agencies, universities, and within industry to give excellent technological guidance for the NREN. I do wish, however, to summarize certain facets of a multi-gigabit network here in order to set the framework for my arguments.

Elements of the new multi-gigabit NREN are:

A. Multi-gigabit multiplexing and transmission.

B. Slow speed switching (or patching) for trunk reconfiguration and special testbed setup.

C. High speed routing and switching (packet switching).

D. Network Interfaces - physical and logical network connection standards, hardware and software.

E. Networking - addressing, routing, switching and control protocols.

F. Applications - Software and data bases which perform specific computational function (providing data access, storage, manipulation, and problem solving).

Transport

We do not need a revolution in the area of transport. Adding new opto-electronic equipment onto existing commercial fiber optic networks can provide multi-gigabit multiplexing and transport within the needed time frame. We can achieve 2.4 gigabits per-second on a single pair within a year, and using new multiplexing techniques, can double that soon after.

Slow Speed Switching

If multiple "T3s" can be used between supercomputers to achieve hundreds of megabits or gigabit rates, then today's technology will suffice for initial gigabit networking. However, new gigabit digital access and crossconnect systems (DACS) need to be developed within two or three years for full network implementation.

High Speed Switching

I am told that today there are prototype classes of packet switches, capable of achieving 5-plus gigibits per second throughput, just emerging from the labs after more than five years of research. Capable of simultaneously switching voice, video, and data, this class of switches will use the Asynchronous Transfer Mechanism (ATM) technology and will ultimately replace our local exchange telephone company central offices and long distance tandem switches. Final implementation issues are being solved as we speak, with deployment anticipated to begin within three years. These new switches with appropriate hardware and software modification can be used as the new high speed NREN switches, thus allowing full multi-gigabit NREN implementation within five years.

We can economically implement these new switches by placing them in parallel, co-located with existing switches. As new upgrades emerge, the new ATM switches can become an integral part of new systems, and we can eventually remove the older equipment altogether . This method allows a smooth economically sound, evolutionary path for the deployment of hardware for a multi-gigabit research and education network.

Network Interfaces

Good low speed interfaces and standards are available today. New 45 megabit and gigabit protocols will need to be developed which contain additional addressing, routing, and control functions. However, connections exist today on a test-bed basis to provide multi-supercomputer applications development.

Networking

This area will require the most work to accomplish a truly flexible multi-gigabit network which is also highly reliable. New addressing, routing, converting, and control protocols must be developed and separated into carefully crafted subsets.

While this work is underway, multi-T3 level circuits can be established over existing commercial fiber based faciltiies to test protocols and get applications underway among major supercomputer facilities.

Applications

The applications are many and will require the collective efforts of hundreds of thousands of researchers and users working independently and in teams, both government and commercial. Some applications will need federal support and vision. Most, however, are best left to the imagination of users and researchers in all walks of life. When the network is available, there will be no scarcity of ideas which can be translated into applications that will drive America to greater heights of success in this new information age. The new NREN corporation must take a leadership role in defining and exploiting application possibilities. We must not wait for users to ask for this help. We must actively seek to help them, and in so doing, achieve our stated mission.

4. Summary

Mr. Chairman, it is possible that a network such as the NREN would evolve on its own, but most likely in one or two decades. It is imperative that the Federal Government take a leadership role, through both policy and funding, to ensure that the NREN is realized in a short time frame on a well coordinated basis. The role of the Federal Government should be one of a partner (both formal and informal) with industry, education, and research. Seed money must be provided by the Federal Government to stimulate new areas of interest and develop them for the greater good of the nation as a whole. It is beyond prudent business planning for corporations to expend these kinds of dollars on their own without a guiding national plan. However, by contributing our business planning and technical resources in an alliance between government and industry, we can greatly accelerate the horizons of research and education in this country.

Furthermore, as it relates to industry, S 1067 calls for the NREN to be phased out when commercial networks can meet the needs of American researchers. Leading edge technologies and applications are always conducted by this team approach of which I speak. Certainly, I believe that the commercial carriers of this country can and should offer the types of services that will grow out of the NREN when it is a sound business decision to do so. This means that there must clearly be sufficient opportunity for return on investment. Also, we must keep in mind the reasons that led us down this path to begin with. The need to be competitive and in the forefront of worldwide research and education must drive us to the commercialization of these services in a short period of time.

For these reasons, Mr. Chairman, I urge that a short but realizable deadline of five years be set for full implementation of the multi-gigabit NREN, and that handoff for commercial operation be required by the end of the seventh year. A slower pace is simply not necessary. Americans have proven time and time again how to accomplish great things in short time intervals when given a clear goal, a new clean structure from which to operate, reasonable funding, and the minimum of government control and red tape.

Open, fair competition will allow the United States to best perform in the global information community. We need the active participation of all players, and through well-crafted, intensely scrutinized procurements. If the Government encourages industry to play hard but fair, the real winners will be the American people.

No one entity must be allowed to dominate this or any other undertaking. Market shares or controls, whether in supply of goods or services, of more than 30 to 40% by a single entity invariably lead to empire building or predatory action in the short term, with higher costs and poorer service long term. The government must make sure that risk and reward are spread in a balanced fashion to ensure the best life for all Americans on a long term basis. This principle should be ensured throughout the development of the NREN.

Mr. Chairman, I applaud you and the many others who have worked so diligently toward the timely establishment of a very advanced National Research and Education Network. I thank you for the opportunity to appear before this hearing today.

Senator **Pressler**. If I could just interject. And I apologize, I have to go and portray myself as an expert on child care on the Senate floor at this moment. But when Mr. Collins was talking, he was talking about small business. And for the record, if you could submit later, what would be the impact if this technology were available to small businesses?

Indeed, this would mean that a mall business that is not near a big city would have a chance to compete—I mean, or not near a great university or not near computers—if this could be available, if supercomputer technology could be available at a low rate to small businesses, would not this mean that a small business located out far away from a big city or away from a great university would gain an advantage?

Would that be a correct assumption?

But we have one more to hear from. But I would love to hear some analysis of that.

And has there been any study of the impact done strictly on small business: what impact this legislation—or the effect this would have on them?

Senator **Gore**. Anybody want to address that before Mr. Schwantes speaks?

Mr. **Collins**. We have taken a look at the kinds of services that would be used by the American at the year 2000. Many of those would be available to the small businesses. And I believe your point is exactly right: we must not isolate the small business in the rural part of the country from being a viable entity in the high-tech world of the future. We need, certainly, the chance to develop the technologies and drive the costs down for bringing that service out there.

And I believe if we go ahead with this kind of network, we will be able to build the volumes and drive the costs down, which will make it economically attractive to rural America. We would like to prime the pump by being able to be involved in the development of those services.

And I would be happy to provide more information at a later date.

Senator **Gore**. I might just say that with these new work stations that are now readily available and are getting cheaper all the time, small businesses in small towns will benefit even if they do not have ready access to the big trunk lines. They will benefit by having access to flows of data and new services that spin off the network, through the work station they will be able to contract out inventory control, do strategic planning in fundamentally new ways, and a lot of other things that are difficult to imagine now. That seems awfully reasonable to anticipate.

Anyway. Mr. Roger Schwantes, Vice President for Product and Technology Planning at Northern Telecom. And as Mr. Liebhaber said, location is very important, and Nashville is an excellent location. So we are delighted to have you here today.

STATEMENT OF ROGER SCHWANTES, VICE PRESIDENT FOR PRODUCT AND TECHNOLOGY PLANNING, NORTHERN TELECOM, INC.

Mr. Schwantes. Thank you, Senator. I am pleased that you noted that we are headquartered in Nashville.

Let me begin by stating Northern Telecom's unequivocal support for the legislation's objectives and for reasons that are included in my following remarks.

Number one, it drives the deployment of a high-speed networking infrastructure, which places supercomputing, scientific and information resources within the reach of all. And that gets at the small business issue that you were talking about before.

It facilitates collaboration and cooperation between geographically dispersed teams of specialists addressing complex issues and opportunities. Once again, that is a resource that small business can tap into. It makes available to small, innovative businesses computer modeling and simulation tools, previously unavailable or inaccessible. It makes available to all, access to the best minds and facilities in the Nation's research community. It provides an infrastructure for translating creative, innovative and market-driven concepts into prototypes prior to final productization.

By facilitating new and improved technologies, processes and competencies, the legislation takes significant steps toward the restoration of manufacturing as the corner of the U.S. economy, which is a serious problem our Nation faces today.

Members of my staff and Bell Northern Research, our research and development affiliate, have developed several recommendations for this committee's consideration. And since you have given your permission to submit that, we will submit that for the record.

For the balance of my time I would like to focus on a single subject: the most effective deployment of the National Research and Education Network. The National Research and Education Network's implementation as described in the Federal Research Internet Coordinating Committee Program Plan, is technologically feasible. The phase three goals of 3 gigabit speeds is consistent and aligned with the planned advances in transmission and switching systems technology.

Northern Telecom and other telecommunications vendors, as we have already heard today, are engaged in development programs leading to transmission equipment operating at speeds of up to 2.4 gigabits; associated exploratory programs investigating methods of supporting up to 9.6 gigabits on an optical line system; and a operating broad band packet switch at gigabit per second rates leading to switch architectures with capacities in the 1-terabit-per-second range.

Exploratory programs in the areas of very high speed logic, electronics, opto-electronic transducers and hybrid integrated circuits, supporting software and control structures required for the design, operation and management of broad band networks; as has already been stated here today, this is a very serious issue, and needs to be addressed.

Standards like SONET, the Synchronous Optical Network, to facilitate broad band network configurations by multiple vendors. As it has also been stated today, if we are going to have an inter-operable network, the standards are extremely important, and we believe that your efforts in this area can be very instrumental in making that happen.

It is clear that there is a universal industry support for broad band networks, and that the speeds and functionality will be even greater than that envisioned in your legislation. Clearly, then, the issue is how to accelerate the earliest deployment and availability of the network so as to gain the benefits that we have talked about.

The National High-Speed Computing Technology Act can play a critical role in resolving this issue.

First, the bill provides a clear direction for the entire industry, thus reducing up-front risks in an emerging market, and providing a goal on which to focus. The governmental role in guiding and supporting the National Research and Education Network's implementation parallels our experience with the space program and the highway system that we have all alluded to.

Second, the bill drives government and industry toward commercialization of this network at the earliest possible date. In order to achieve this, it is imperative that there be no single-vendor solution; that the entire industry be allowed to engage in the free competition of concept and ideas.

To ensure ubiquity of access, this must include unleashing the power of all local and inter-exchange carriers without the more cumbersome aspects of the regulatory process. That has been commented on earlier today, also.

The National Research and Education Network is a national imperative. As such, government support and guidance is necessary to achieve the goal.

Third, the bill's commitment to early deployment will drive industry cooperation on national and international standards, thus permitting U.S. leadership in an emerging international marketplace.

Fourth, the bill will provide financial support to leading-edge applications and pilots to ensure commercialization of the network in the shortest possible time.

And fifth and finally, the bill should incent and provide for alternate uses of the multi-media high band width capability the network offers, thus achieving a critical mass of traffic volume and the resultant economics of scale.

In conclusion, we are obviously excited about the prospects and opportunities and benefits that the National High-Speed Computing Technology Act provides for all segments of the population.

We appreciate, Mr. Chairman, the process of consultation with industry by which this legislation was fashioned.

And I will be happy to answer any questions at this time.

Senator GORE. Great. Thank you very much.

Do any of the other witnesses have any comment on Mr. Gabbard's suggestion about creating an independent corporation that would phase itself out, but would be separate and distinct from NSF as a mechanism for managing and ram-rodding this project?

Mr. Liebhaber.

Mr. LIEBHABER. May I make a comment?

Senator GORE. Yes.

Mr. LIEBHABER. Senator, and again, I am just speaking from the experience of our relationship with the National Science Foundation in the current NSFNET project. We are very encouraged and very optimistic about the work the National Science Foundation has been able to do in bringing together the inter-agency cooperation for the second phase of the National Science Foundation Network.

I would suggest that there really are so many special interest issues here that a very independent organization such as NSF, the National Science Foundation, really appears very appropriate to us.

I would like to answer another question that you asked the first panel, which was what else is it this committee might do.

To move this infrastructure along. I would suggest that the most difficult problem we foresee in this area of development is the capability to access wideband and digital facilities in the last and/or first mile. And anything this committee can do to encourage the local exchange carrier to modernize facilities would be very beneficial.

Senator GORE. Well, we are trying to do that in a separate measure. I am proposing to free up some of the restrictions on participation in the cable industry. Entertainment, historically, has driven that kind of first mile/last mile connection. But that is another whole separate subject, with a controversy all its own.

But, back to the first question: you think NSF does have the requisite expertise and independence.

Anybody else want to comment on the issue.

Mr. Collins.

Mr. COLLINS. I like the notion that there would be more direct business involvement in the managing and setting the time frames, objectives and melding of the ideas. The key here is the drive behind it, and there is nothing like profit motives to cause that drive to go quickly. And there is a huge integrator function. You need participants from all segments of the industry.

Senator GORE. How would you go about that?

Mr. COLLINS. Well, the non-profit organization to get it started and to provide the shield for the people to work together in the diverse segments of the industry, sounds like, on the surface, not a bad idea.

Senator GORE. Well, what if NSF contracted it out? If NSF provided the supervisory function using the expertise that exists there, but then contracted out the management role? What about that?

Mr. COLLINS. Well, in my comments, I have suggested that one of the ways to do this is to put the frame, whatever it is that we want in the capacity of this network, how wide we want it, how fast we want it to grow, and put it out for bid and let consortiums come together and make bids on providing this kind of network. A technology will have to be developed; you will have to have a lot of players in it, but if you look at the old——

Senator GORE. In other words, include the management function in the solicitation for bid?

Mr. COLLINS. That would be one way of doing it.

Senator GORE. But maybe keep the supervisory role in NSF?

Mr. COLLINS. Somebody has got to set the goals and standards for the network.

Senator GORE. Right. You know, the tendency has been to set up a new entity for stuff like this, but maybe that is not called for here. But I am intrigued by that wrinkle on Mr. Gabbard's suggestion.

Anybody else want to comment on it.

Mr. Schwantes.

Mr. SCHWANTES. Yes, I share some of the same concerns about having a multi-interest group trying to manage something like this. And I think that we would favor the National Science Foundation to at least supervise it and establish the standards for the network and that the business relationship issues probably be handled in the way that my colleague has just mentioned: in fact that this probably will have to go to a bidding process.

Our concern in the whole process is that we may end up leaving some major players out of it. I do not have a good solution for you at the moment, but I think we would like to give it some more thought and resubmit some information.

Senator GORE. Very good.

What about the question of protocols? Does anybody have any additional thoughts on protocols? Is that really a bottleneck?

Mr. SCHWANTES. Protocols, without a doubt, standard setting is the major bottleneck in the industry. I think we are all ready to go forward on broad band networking, but to have a ubiquitous interoperable network, the standards issues need to be addressed up front.

One way that that can happen is if we move to the trial state as quickly as we can to demonstrate that these networks can operate in the fashion that we want them to.

Dr. LUCKY. Mr. Senator, if I might comment on that. I think protocols are an example of one of the research issues that the bill has to attack, because not only do we need the standardization and agreement on that, but the protocols that we have today do not do the job for a gigabit network and there are a number of issues like that where we really need to get our hands on these streams and that is what is good about your bill. You put together a community of researchers who try to actually use this and today you cannot stuff that many bits through a protocol. It just does not work that fast.

Senator GORE. Well, 100 years ago in Europe every railroad used a different gauge track and there were no interconnections and we want to avoid a situation like that.

I am fond of quoting Yogi Berra, who once said "What we have here is an insurmountable opportunity," and the three gigabit network is an opportunity we must make certain does not fall in that category.

But protocols, as the FCCSET pointed out as well, represent one of the key bottlenecks that we have got to address.

Mr. Gabbard?

Mr. GABBARD. I really would like to go back to the previous question for a second, if I may, and that is on a procurement cycle, because I believe if you put out one massive procurement cycle like

FTS-2000, that one could consume two to three years just in that process and end up with a sub-optimum solution.

I think the best thing to do is to either task the NSF, if that is the chosen body, or set up a new organization and let them get underway quickly putting out bids first for transmission.

As soon as they have standards for switches, for example through DARPA, let DARPA manage the procurement of the switching requirements. Get the National Institute of Standards and Technology and, for example, the IEEE working on the standards right away and get them funded right away, so that you can have hundreds of smaller procurements that can be done quickly and can be managed efficiently.

I really propose that as the method of moving this thing within a reasonable time frame and without a tremendous infrastructure. It is not needed if you do it that way.

Thank you, sir.

Senator GORE. Does anybody think that does not make sense?

Mr. SCHWANTES. It makes sense to me.

Dr. LUCKY. Do not forget the systems integration is probably the toughest problem here, so I do not mind the notion of having a lot of little contracts but somebody has got to do a lot of work to put it all together, and that somebody has got to be very good.

Mr. COLLINS. I do not believe, having that many diverse, unrelated organizations working in separate pieces is the right way to solve the problem. I think if you put it together as the total network bid, the parties will make all the pieces fit together. I think that would take four years rather than three years.

Senator GORE. Well, we do not quite have unanimity on this particular point, but we will try to sort that out. We have had a long day and I have got a lot of other questions, but we really do need to wrap it up.

You all are invited to expand your comments for the record, but at this point I want to conclude by thanking all of our witnesses.

I think it has been an unusually productive hearing, the kind of hearing that generates significant momentum behind an idea that needs to be implemented. The statements have all been very impressive today, and I am grateful to all of the witnesses.

I might say that these discussions are going to continue. We will have some other hearings, but we are going to try to move this legislation. The ranking Republican member of the Subcommittee, Senator Pressler, has joined as a co-sponsor. Others have joined as co-sponsors as well. The bill is picking up momentum.

I testified yesterday on the House side, where many members expressed a great deal of support. In fact, the House Science & Technology Committee has been the partner of this Committee from the very beginning. Some of the key ideas have originated on the House side with the Committee and with members of the staff of that Committee and there is every indication that they will soon be considering a parallel measure.

So I think that for those who are interested in developments of this sort, the prospects for this actually happening are improving rather rapidly.

In any event, we are going to have some other discussions later this evening with some people staying in town to talk about some of the issues involved and we will have other hearings that we will be scheduling soon.

With that, let me thank everyone again and declare this hearing adjourned.

[Whereupon at 4:55 p.m. the hearing was adjourned.]

NATIONAL HIGH-PERFORMANCE COMPUTER TECHNOLOGY ACT OF 1989

WEDNESDAY, JULY 26, 1989

U.S. Senate,
Committee on Commerce, Science, and Transportation,
Subcommittee on Science, Technology, and Space,
Washington, DC.

The subcommittee met, pursuant to notice, at 1:32 p.m., in room SR-253, Russell Senate Office Building, Hon. Albert Gore, Jr. (chairman of the subcommittee) presiding.

OPENING STATEMENT BY SENATOR GORE

Senator Gore. The subcommittee will come to order. I would like to welcome all of our witnesses and guests.

This afternoon the Science Subcommittee will examine some of the latest developments in computer software and see demonstrations of supercomputer graphics and ultrasophisticated computer models, and we will hear about computers you can talk to and computers that teach.

There is a lot going on, to say the least, at the leading edge of computing: software for supercomputers and other leading edge computer systems, artificial intelligence programs, new programming languages to make programming easier, and technologies that make computers easier to use.

Today we will look at the potential benefits of advanced computer software and the challenges we face in trying to realize those benefits.

Visualization and artificial intelligence are particularly exciting. Visualization allows scientists to use supercomputers to create stunning computer graphics—pictures that make mountains of numerical data instantly comprehensible.

Supercomputers can generate hundreds of millions of bits of information every second. If that much data were typed out as a list of numbers, it would be several miles long, and of course no one could make any sense of it at all.

Visualization recognizes that far more than half of the human brain is devoted to processing images and that for us visual images are by far the most efficient way to understand and communicate information.

We have all heard the old saying, "A picture is worth a thousand words." Well, with visualization it is more like a picture is worth a billion bits.

The work being done on visualization today will soon be applied outside the laboratory. Computer technology and networking technology are developing so quickly that the supercomputer graphics we will see today we will be able to see at home in only a few years.

At one recent hearing, John Rollwagen, the CEO of Cray Research, described how the personal computer on his desk is just about as powerful as the first Cray supercomputer that his company installed 13 years ago.

What is at the leading edge today will be ordinary and mundane by the year 2000. That is why research on visualization and advanced software is so important. There has been a great deal of talk about HDTV and how the Japanese are going to beat us in that market, but Japanese HDTV is just a better version of old analog TV technology.

We cannot just worry about HDTV. We need to pay attention to the next revolution in TV technology, digital TV.

The technology being developed for visualization can be used to develop super-high resolution digital TV, and the same fiber optic network that today links computers could be used to carry digital TV signals in the future.

The United States has the lead in visualization and networking, but we need to redouble our efforts in order to keep that lead and to leapfrog the Japanese in several areas of technology. We can lead the digital TV revolution.

Artificial intelligence is also at the leading edge and holds exceptional promise. This software allows computers to mimic the way people think, to take input from the outside world and make logical conclusions.

For example, NASA is developing vehicles to explore the surface of Mars which can in a sense think for themselves. These vehicles would use cameras and radar to detect obstacles. Then AI software would interpret the camera and radar images and determine a safe course around rocks, craters, and other hazards.

Artificial intelligence is finding applications in the classroom as well. Students can learn from computer tutors, computers that respond to the student's individual learning needs. Such one-on-one computer teaching can keep students interested in learning and, since computers never get tired or impatient, they might be able to relieve some of the pressure on overburdened teachers.

Some of the most important applications of AI are in programming. Modern computer programs can be incredibly complicated. Computer programs containing a million lines of code are becoming commonplace in the military now.

Ensuring that such programs are error-free is a daunting task, to say the least. So computer scientists are beginning to use AI programs to find the bugs in their programs and in some cases to write their programs.

Such programs are part of the solution to the so-called software bottleneck which currently slows development of new computer systems. Computer hardware is improving so quickly and getting so complex that the software just cannot keep pace.

It is taking more and more time and effort to build the new software needed for a new computer system. In 1960 if you bought a computer system, 80 percent of the cost was for the hardware—the chips, the wiring, the disk drives and so forth.

Today it is just the opposite: 80 percent of the cost of a typical computer system goes for software development and maintenance. And this trend is going to continue with our renewed effort to make software development easier and more efficient.

We cannot take full advantage of the promise of visualization and artificial intelligence without paying more attention to software engineering.

Throughout today's hearing, we will be asking the question, how can the Federal government promote development and use of advanced computer software. The Federal government and especially the military is the country's largest consumer of software.

In recent years there have been hundreds of weapons systems crippled by bad software. Good computer hardware, of course, is next to useless without good software. So clearly we would save taxpayers' money with better software that is easier to use, easier to write, and less prone to errors. We need more R&D so advances in visualization and AI can be put to the best use.

In May I introduced S. 1067, the High Performance Computer Technology Act, in order to roughly double the amount of funding for development and use of advanced computing.

Today's witnesses will clearly demonstrate why this is important. There are incredible benefits to be gained in this field if we make the necessary investment.

Several members of this subcommittee have joined me in sponsoring this legislation, and I hope that others will support the effort as well. It is an ambitious plan which authorizes $1.75 billion in new funding over the next five years.

Budgets are tight and it will not be easy to find the necessary funding. But this is new technology and we need new money to develop it. Our economic competitiveness and our national security depend on it.

S. 1067 also contains several measures to promote advanced computing. Most important is the three gigabit national computer network, which would link supercomputers and researchers around the country.

Just last Friday, the President's new Science Adviser appeared before the Commerce Committee and said that such a network would be the single most effective way to improve the productivity of American researchers. It would speed up the development and sharing of computer software and provide supercomputer access to millions of researchers around the country.

There are a number of other provisions in the same legislation to promote software development. Almost half the funds authorized by the bill are authorized for advanced computer software.

In addition, the National Science Foundation is required to set up software clearinghouses to allow researchers to find the research software they need. The National Institute of Standards and Technology is required to develop software standards to make it easier for software that runs on one computer system to be transferred to a different system.

So today we will look at other ways to help break the software bottleneck and to take advantage of new developments in visualization and artificial intelligence. Computer technology is advancing at breakneck speed. Policymakers must be sure that government policies and research priorities do not lag too far behind.

Senator Robb.

Senator Robb. Mr. Chairman, I have no opening statement. In fact, I am unfortunately not going to be able to remain here for the entire hearing. I have to go preside.

But I am going to take the statements with me, I look forward to returning if the hearing is still taking place.

Senator Gore. Thank you very much.

I do have a statement that the Chairman, Senator Hollings would like to have included in the record.

[The statement follows:]

Opening Statement by the Chairman

Today's hearing will focus on an aspect of computer technology that often gets less attention than it should. You can't see software, and you can't touch it. It is much easier to talk about something tangible, like the hardware that makes up a computer. Yet the software is critically important, and I am glad Senator Gore decided to call this hearing to take a closer look at the benefits of and the problems with advanced computer software.

As someone who is concerned about making sure the U.S. has a strong defense, I pay a lot of attention each year to Defense Department programs. And I have heard time and again about software problems that have delayed the development of weapons systems. As the systems get more complex, the bugs in the software have more places to hide and are more expensive to find and fix. Sometimes it takes tens of millions of dollars and several years to get things working the way they are supposed to. The B-1B avionics system is one of the most glaring examples.

Research on software is critical if we are going to streamline software development and stop wasting taxpayers' dollars. Many of the innovations to be discussed today need to be put to good use quickly in the private sector and in the military. Innovation alone is not enough; we must put this new technology to good use.

I look forward to reviewing the testimony from this hearing and learning about the new advances being made in computer software. In addition, I look forward in coming months and years to helping see that innovative software gets transferred from the laboratory to the offices and the factories where it is needed. Our competitiveness depends on it.

Senator Gore. Our first panel consists of two men who know a great deal about software and about the Federal government's role in software R&D. Let me invite them up to the table at this time. Dr. William Wulf and Dr. Jack Schwartz.

Dr. William Wulf runs the Directorate for Computer and Information Science and Engineering at the National Science Foundation. Dr. Wulf is a software engineer, and before coming to NSF in 1988 was a professor at the University of Virginia.

Like Dr. Wulf, Dr. Jack Schwartz is on leave from academia and is now Director of the Information Science and Technology Office at the Defense Advanced Research Projects Agency, which oversees much of the Defense Department's basic research on computers and software.

DARPA has funded much of the pioneering work in artificial intelligence and other areas of advanced computing.

Dr. Wulf, we will start with you. I will say to all our witnesses today that, without objection, the prepared statements will appear in full in the record and you are invited to summarize your formal presentations. And as I mentioned, we will start with you. Welcome and please proceed.

STATEMENT OF DR. WILLIAM A. WULF, ASSISTANT DIRECTOR FOR COMPUTER AND INFORMATION SCIENCE AND ENGINEERING, NATIONAL SCIENCE FOUNDATION

Dr. WULF. Well, thank you very much, Senator. As I listened to your opening statement, I was wondering whether I could add anything to it. It was a marvelous statement of the situation.

Senator GORE. It was not that long, was it?

Dr. WULF. No, sir.

I think you hit the nail on the head in that you properly characterized the software problem as our inability to produce software, to produce it cost effectively, to produce it on time, and to produce software that operates reliable, and that is sufficiently useable by the intended user community.

I suppose one could say, so what? You know, everything is expensive and takes longer than we want, and so what is the difference? This is like a lot of other things.

I do not think the general population appreciates how important software is to both our economic and our military health.

Let me focus a moment on the economic issues, and Jack can deal much better than I with the military issues. By one estimate, the software industry was a $50 billion industry in 1988, and estimated to be a trillion dollar industry in the year 2000.

I think that badly understates the importance of software to the country. More and more software is embedded in commercial products and embedded in a way that is invisible to the average consumer. I tried to give examples in my written testimony in products ranging from wristwatches to airplanes. I was struck, however, by another example which I would like to share with you.

It turns out that my executive officer owns a Porsche automobile and gets the Porsche Sports Car Club of America magazine. In the July 1989 issue on the inside front cover, there is a full-page ad. The ad is for a chip that you can buy that will boost the horsepower of your Porsche 10 to 15 percent. In this case at least, the chip is merely an electronic recording of software; it is the way the software is delivered, much as you would with a disk.

The moral is, if you want to sell high performance automobiles you have to be able to produce software. And the same thing is true if you want to produce microwave ovens or VCR's or HDTV's or virtually any consumer product as we go out into the next decade and the next century.

In addition, of course, software is critical to the manufacture of products, everything from the design in computer-aided design software through computer-aided manufacturing software.

But there is another kind of software that is, I think, of special importance and deserves special mention here, and that is scientific software. As we have discussed in previous hearings, computational science, the application of computers and information processing technology to scientific and engineering investigations, is becoming increasingly important.

We refer to it now as computational science and engineering. Of course, science and computers have had a long history. Some of the earliest uses of computers were for scientific investigations.

But it has gotten to the point where computational science is really a new modality of scientific investigation, complementing the observational, experimental, and theoretical modalities of science.

If we want to compete effectively in the scientific arena, where we have prided ourselves in the past, we are going to have to be able to produce the kinds of software which will support those scientific investigations.

Too often I see disciplinary scientists—biologists, chemists, physicists, and so on—devoting a large fraction of their professional life to developing software rather than to their basic science, simply because the software that they need is not available and there is not a large enough market for there to be a significant number of vendors from whom they can procure it.

Let me talk for just a minute about the cause of the software problem. It really is a rock and a hard place that we are between.

As you commented in your opening statement, software is extremely complex. Some of the modern software systems are among the most complex creations of the human mind. I was searching for some way to characterize that to the people who are not directly involved in software and ran across the following analysis. It was done about 10 years ago on the FAA en-route air traffic control system, which at that time contained about 600,000 lines of code. That by today's standards would be considered to be only a moderately large system.

By at least one estimate, the number of paths—that is, the number of sequences of instructions which could be executed through that program—is about 10 to the 11,800. That is 1 followed by 11,800 zeroes. Every one of those must be correct in order for the software to work correctly.

How big is 10 to the 11,800? Well, if you want a benchmark, and if I remember my freshman—or maybe it was my junior—physics correctly, it is estimated that there are about 10 to the 120 atoms in the universe. I do not mean on the Earth or the solar system or our galaxy; I mean the entire universe.

The number of paths in the FAA program is therefore 10 to the 11,680 times larger.

Okay, so the rock of the software problem is the fact that software is complex. The hard place is that the construction of software is a craft industry. And I do not mean that in a pejorative sense at all. It is a very difficult problem.

It is a very intellectually demanding problem. It requires great mathematical ability. But nonetheless, absolutely every characteristic of a software product depends upon the craftsmanship, the ingenuity, the knowledge of the people constructing it.

So here we are between a rock and a hard place. The programs we construct are effectively too large for humans to understand, and yet every characteristic of them depends upon the human's ability to understand them, to cope with them.

All right, so that is the rock and the hard place. That is what causes the problem. What is the solution?

In some sense, we have known the solution for a long time, and it is really implicit in the statement that I made before, and that is we have to make software at least apparently less complex, we have to build a human resource base that can construct it, and we have to automate more of the craft aspects of the development of software.

The problem, of course, is knowing how to do it, how to do any of the three things I just listed. I should point out that the fact that software is a problem is not new news. In fact, there was a workshop held in Garmisch, West Germany, in 1968, 21 years ago, which declared "the software crisis."

For 21 years we have used the term "software crisis" rather liberally. We have tried lots of solutions. Lots of ideas have been proposed to solve the problem. We have made progress, but we have not solved the problem.

It seems to me that it is time to recognize what is probably the obvious, and that is that it is a very hard problem. It is a problem for which we are not going to find a quick and easy solution.

It is a problem which will require a commitment to long-term basic research. I do not believe that we even have the appropriate mathematical basis for characterizing software and, lacking that, we will not develop the appropriate engineering base either.

The National Science Foundation has of course had many activities in the software area for many years. Since coming to the Foundation a little over a year ago, I tried to make it one of my highest priorities, and we are starting to make some progress, I think.

Two of the science and technology centers that were established by the Foundation the early part of this year bear directly on the software problem. One is centered at Rice University in Houston and is focused on the problems of parallel computation, one of the most important problems, I might add, for scientific computing.

The other is centered at Rutgers University in New Jersey and is focused on theoretical computer science and its relationship to discrete mathematics. Here again, I think much of the fundamental software problem will hinge on developing a better understanding of the theoretical aspects of the problem.

I think these are only a start. There are some short-term things that we can and should do. I assume that some of the other people testifying will talk about them, and others are contained in my written testimony.

But the basic message that I really want to convey to you is that it is a very hard problem, a very important problem, and we need to make a commitment to long-term basic research to solve the problem.

Thank you.

[The statement follows:]

Testimony

Dr. William Wulf
Assistant Director
Computer and Information Science and Engineering
National Science Foundation

Thank you Senator. I appreciate the opportunity to testify today on this important issue.

Let me begin by saying a few words about myself, and hence the perspective that I bring to this issue. I am the Assistant Director of the National Science Foundation with responsibility for the Directorate of Computer and Information Science and Engineering (CISE); CISE funds about half of the Federal share of basic research on software. I am on leave from the University of Virginia, where I am a member of the Department of Computer Science, and the AT&T Professor of Engineering and Applied Science.

Prior to joining the UVa faculty, I founded and was the CEO of Tartan Laboratories, a software company specializing in compilers for Ada -- the new standard DoD programming language for embedded systems. The technical basis for Tartan was research that I had done while a Professor of Computer Science at Carnegie-Mellon University. I have served as a consultant on software and computer architecture to many major computer companies and the government. I come from a family of Computer Scientists; my wife, Anita Jones, is the Head of the Computer Science Department at UVa, is a member of the Defense Science Board, and was on the DBS Task Force that wrote the 1987 report "Military Software".

So, as you see, I could comment on the problem from a number of perspectives. However, in the interest of time I will confine my remarks primarily to issues of software research.

Importance of Software

Let me begin with the enormous importance of software to the nation. By one estimate, software was a 50 billion dollar industry in 1988, and will become a trillion dollar industry internationally by the year 2000. Although large, I believe that both these numbers underestimate the crucial role of software to the economy and to our ability as a nation to compete in the world marketplace.

Software is what makes computers useful, of course, but it much more than that. Increasingly software is invisibly "embedded" in products as diverse as wristwatches, cameras, telephones, automobiles, and aircraft. The software in these products is crucial to their functionality, and hence to their competitiveness. Product differentiation in microwave ovens and VCRs, for example, does not come from the quality with which they perform their basic tasks -- they all heat food and/or display videotapes with approximately the same quality. Rather, product differentiation comes from the "functions" they provide, all of which are provided by software. The revenue figures cited above do not include that of cameras, VCRs or automobiles -- yet, they could not be sold without their software.

Software is critical to manufacturing: Computer Aided Design (CAD) is crucial both for creating superior designs and for producing them in a competitive time-frames; Computer Aided Manufacturing (CAM) and robotics software is crucial to both cost-effective and high-quality manufacturing; inventory control and automated order processing software are critical to economical methods of production, such as "just in time" manufacturing.

Software will control the new generation of high-performace, fuel-efficient, "fly-by-wire" aircraft. Software routes our phone calls over the telecommunications network, dynamically adapting to both load and malfunctions. Software controls the bar-code readers at our supermarkets, ensuring correct billing of customers and providing continuously updated inventory control. Software runs our elevators and provides environmental control in our office buildings. Software controls automatic teller machines, providing us 24-hour banking services. The list is virtually endless.

There is another kind of software of strategic importance and which must be of national concern -- scientific software. The use of information processing technology is emerging as a new, "fourth modality" of scientific investigation. Supplementing the observational, experimental, and theoretical modalities, "computational science" is an enabling technology permitting a whole new class of scientific and engineering applications, including the design of new materials, the design of new drugs, and the understanding of both macroscopic and microscopic phenomena inaccessible in any other way. Whoever masters the production of scientific software will have a huge head start in tomorrow's economy.

There is a tendency to think that software is the exclusive concern of the computer industry; that is wrong! There is virtually no aspect of our economic or military posture that is not critically dependent on software. Software is a problem of major importance to the entire nation.

What is the Software Problem?

The production of software is both very expensive and unpredictable. It is often delivered over budget and behind schedule. When it is delivered, it almost always incorrect, and the cost of fixing the errors may be many times that of the original development. Even when reasonably correct, initial versions of software are typically hard to use, lack important functionality, and are much less useful than had been originally intended. Changes to operational software, to respond to advances in technology for example, are typically even more costly and error prone.

Recognition that software is a major problem is not new; in 1968 NATO sponsored a workshop in Garmish, West Germany to discuss it, and at that meeting the term "software crisis" was coined. To be sure, significant progress has been made in the intervening 20 years, but so has our need for high quality software. Indeed, our need for software has increased much more rapidly than our ability to produce it. The best documented evidence of this comes from the avionics industry, where the demand for software has increased 20% per year while productivity has increased less than 7% per year. The difference has been made up by increasing the number of people, but that cannot continue indefinitely.

The cause of the software problem is easy to state; it arises from two simple facts: (1) the complexity of software systems, and (2) the "craft" nature of their development.

(1) Large software systems are among the most complex creations of the human mind; many of which have exceeded the ability of humans to comprehend them.

Software is composed of individual "statements" that instruct the computer how to transform data. For a software system to function properly, each such statement must (a) work properly on billions of different data values, and (b) interact properly with every other statement in the system. This second requirement causes the complexity of a software system to increase exponentially with its size. Typically, students struggle to compose correct software containing a few dozen statements. Production software systems, by contrast, may contain millions of statements, and be trillions of times more complex than these student exercises.

(2) Software development is a craft industry. This is not meant to be a perjorative remark; it is an extremely difficult intellectual activity requiring deep mathematical and engineering skills. Rather this statement is meant to convey the dependence of all aspects of software products on the craftsmanship of their creators. Everything from the functionality, size, speed, correctness, and maintainability of the product depends upon the knowledge, creativity and workmanship of its authors.

These are the "rock and a hard place" of the software problem: software is complex beyond the ability of humans to comprehend it, yet its properties all depend on the craftsmanship of those same humans.

What is the Solution?

At least the shape, if not the details, of the solution to the software problem is implicit in the previous statement. We must reduce the (apparent) complexity of software systems to the point that they can be understood; we must build the human infrastructure capable of quality software development; and we must replace labor-intensive, craft aspects of its production with capital-intensive, automated processes.

This much has been known for a long time; alas, knowing how to do it is still a deep intellectual problem that will require a commitment to long-term basic research and human resource development.

I must admit to you that I did not always believe this; at one time, I believed that the problem lay in getting industry to use the research results already available. That is also a problem, and one that I will return to later. But, I am now convinced that it is the lesser of the problems. The much larger ones are that the fundamental intellectual foundation, even the appropriate mathematics, does not exist. Our educational system is struggling to cope with this new discipline; it is nowhere adequate -- not at the graduate level, not in the elementary schools.

Where is the US?

At the moment, the United States is the world leader in the software industry, with about 80% of the total market (50% of which is within the US). We are not, however, appreciably ahead of the rest of the world in basic software research. Japan has clearly enunciated national software research goals, and has initiated well-developed government-industry-university programs to achieve these goals. More significantly, Europe, which historically has taken a more mathematically-oriented, fundamental approach to software research, is beginning to reap the benefits of that approach. Through large government-industry programs, such as ESPRIT, RACE, and ALVEY, which team universities and industry, they are making significant strides toward applying these results.

I would like to point out that one of the usual strengths of the US economy, namely the private sector, is only a small factor in the software research equation. There are no large software companies! Instead, there are a large number of small, young, entrepreneurial companies. We cannot look to them as the source of long-term basic research; they don't have the resources.

Scientific software is a special problem. Commercial and military software have sponsors that ensure that it gets done *somehow*, albeit more slowly and more expensively than one might like. Scientific software has fewer sponsors with deep pockets. Too often, disciplinary scientists (chemists, biologists, ...) with sketchy backgrounds in software development are being diverted from their primary scientific investigations to develop the software they need.

At NSF we are trying to place increasing emphasis on software research. Since coming to the Foundation last year, I have made it one of my highest priorities. There has been some progress, however, and I would like to mention some recent developments:

> Two of the recent Science and Technology Center awards from the Foundation. One went to a group based at Rutgers University to study Discrete Mathematics and Theoretical Computer Science. Another went to a group based at Rice University to study parallel computation -- one of the most difficult of software (and hardware) problems, and one of special importance to scientific computations.
>
> We have teamed with DARPA, the other major Federal supporter of basic software research, and have undertaken initiatives on the Theory of Parallel Computation, and on Artificial Intelligence. We are studying the feasibility of another joint initiative on Formal Methods in Specifications of Software.

We need to build on these beginnings, and to forge stronger links with industry to commercialize the ideas that continue to emerge from academic research. The direction is right, but much more remains to be done.

What Needs to be Done?

The most important action is to continue our commitment to long-term fundamental research. This is perhaps the least glamorous action that one could contemplate, but it is an investment in the future of the country. Such research is the source of ideas and steady progress, and a primary mechanism for expanding the skilled human infrastructure needed for this complex field. As I noted earlier, many of those now working in the software industry were trained in other disciplines and have a weak foundation in Computer Science. The problem is getting worse.

There are shorter-term actions that can and should be undertaken immediately that may have a significant impact while the longer-term actions have a chance to work. The government needs to reexamine: (1) government procurement policy for contracted software development, and (2) intellectual property laws with respect to software.

Current government "rights in data" policy for contracted software development has had the unintended effect of being a strong disincentive to develop or use modern software technology. The policy requires that the contractor deliver, not only the resultant software, but also any tools and technology used to develop it. The rationale for this policy is that it allows the government to maintain and enhance the resulting software after it has been delivered. In practice, however, it makes it foolish for the bidders to invest significantly in modern technology only to have to give it up. Often such investments are simply not made, or, worse, if made, the technology isn't used on government contracts -- which means that the software delivered to the government is of poorer quality than it might be otherwise.

I know of no software developer that believes that either copyright or patent law is appropriate to software, or that either has the intended effect of creating an economic incentive for its production. This is a particularly difficult problem, and one in which I am not an expert. I am convinced, however, that only a completely fresh, start-from-scratch look at the problem will be effective.

Summary

Software is crucial to all aspects of the country's vitality -- from formulation of scientific concepts, through design and manufacturing, to product functionality. Our ability to produce software, however, is severely hampered by an inadequate scientific base and limited human resources. There is no "quick fix" for these problems, and the most important step we can take is to expand our commitment to fundamental research and human resource development.

Senator GORE. Very good. Thank you.

We will hold off on questions until the next witness is finished. Dr. Jack Schwartz has already been introduced formally. Welcome. Please proceed.

STATEMENT OF DR. JACOB T. SCHWARTZ, DIRECTOR, INFORMATION SCIENCE AND TECHNOLOGY OFFICE, DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

Dr. SCHWARTZ. Thank you, Senator.

I will in fact be in good accord with what Bill has said and will pick up on some of his points. Let me say incidentally that software is a problem that I have been thinking and writing about for about 20 years as a university professor and then industrial consultant.

I would like to begin by drawing a contrast between computer hardware and software. Computer hardware has advanced with tremendous and steady rapidity over the last 20 years. Software technology has advanced, but not advanced nearly with the same steadiness or to the same extent.

This represents not an absolute failure, but relative failure of the software area which has turned software into the great unswallowable lump of the whole computer field, so to speak.

Senator GORE. To some extent, the rapid advances in hardware have complicated the evolution of the software.

Dr. SCHWARTZ. That is very true, yes.

Now, I would like to point out that there is a basic technical reason for the difficulties experienced in software: it falls to the software developer to deal with all the residual complexity accumulated by every other area of technology. Whenever a hardware developer comes to a point that is intrinsically complicated, it is in most cases right for him to say: I am not going to solve this; I am going to leave it over for the software developer. This is the line of maximum technical advantage.

The consequence is that the software problem is not one problem. It is the mass of all problems, of all complications inherent in every other design problem, thus forming a kind of tarry residue of complexity.

I would therefore compare it, not to the problem of reaching the Moon as that problem might face NASA, but to the complex of problems that a physician or that the world of physicians faces in prolonging life and adding to health. This also is not one problem; it is a mass of problems.

Everybody understands that when one of these medical problems is solved, it will only uncover the next layer of problems. So that, even though everybody expects the advance of medical science to make people healthier and healthier, as it has, nobody expects immortality to be the near-term result.

The importance of software has been called out in many DoD reports, for example, in the recent report by George Millburn on defense critical technologies, which designated software as one of the 22 defense critical technologies. Let me read a few figures that point out its very essential military role.

The F-16A, in 1981, carried seven physically separate computer systems, comprising roughly 50 individual digital processors. It carried 135,000 lines of code.

Five years later, the F-16D carries six times as many processors: 300 processors in total. Instead of 135 thousand lines, it carries a quarter of a million lines of code. So the complexity of the aircraft software requirement has doubled over those five years.

A look at a system like the B-2 shows far more stupendous numbers. The B-2 will carry over 200 computers and 3.5 million lines of code. In addition to those 3.5 million lines of "flying code," roughly 18 million lines of code are required for various ground support functions.

This should make it apparent that systems like the B-2 simply cannot function without an enormous mass of software.

Let me expand now on the point that you have just made, Senator Gore, to wit that the physical advance of computer technology has provided some assistance for the software developer. An important example of this is that we can now use computer power in a lavish way to replace the human craft labor of the programmer.

On the other hand, hardware development has also complicated the task of the software writer, for example by requiring him to deal with parallel systems in which tens or even hundreds of thousands of simultaneous activities need to be coordinated.

That is a major new combinatorial problem and one on which we are only now beginning to get a grip.

I would like to say a bit about the DARPA programs that address the problem of the software area and then something about the special capabilities of DARPA to function in this area. Since every one of the multiple agencies participating in research on the broad technical challenges of software will bring its own capabilities and twists to bear on the national effort required.

As to programs: we spend about $25 million a year in the DARPA Information Science and Technology Office on software research and tool development. This figure includes specifically network-related system.

In addition to the work managed by ISTO, DARPA runs several other major software efforts and other programs. One is the so-called STARS program: Software Technology for Adaptive, Reliable Systems. This program is currently budgeted at roughly $12 million, and is managed by the DARPA Defense Manufacturing Office, which also manages the Software Engineering Institute at Carnegie-Mellon University.

STARS focuses on the development of a comprehensive set of Ada-based tools for constructing and managing large software systems. These are very large systems, typically systems requiring over 500 person-years in development.

Developments of this size typify many defense software efforts.

The Software Engineering Institute has a technology transfer focus. It aims to find the best and most relevant new developments in software engineering technology and bring them into the defense system world. SEI operates a very extensive program of publications, training courses, seminars, workshops and consulting services.

I do not want to say much about the many programs that ISTO sponsors, but shall review just a few particularly salient examples. We have been involved in the last few years in the development of a very promising new operating system called Mach at Carnegie-Mellon University. Let me try to explain what the advantage of this system is. As you know, operating systems are the packages of software that establish the basic environment of screen, keyboard, communication, printing, and other functional capabilities that you expect to find on a computer when you sit down to use it.

It is very important from the user point of view that the environment available to him on one computer be the same as the environment available to him on another computer. One does not want to have to learn a new working environment, a new set of editors, a new set of conventions, a new everything, when one has to switch computers. You do not want to have to adapt all your applications software to a different environment.

For that reason, portability of operating systems, i.e. the ability to run in identical form on multiple computers, is crucial. Portability has accordingly been much emphasized in the design of the so-called Mach operating system. Mach has been quite successful in this regard. It is running now on close to two dozen different important computers. One of them is the NeXT work station developed by Steve Jobs organization; another is the parallel machine developed by Encore Computer Corporation. I note in this latter connection Mach was also designed from the beginning to serve the operating requirements of parallel computers.

We, along with NSF, are participants in the Rice University work on parallel software tools. We are also in the process of launching a major new activity aimed directly at the general simplification of software design. Let me explain a bit about this important new effort: our so-called common prototyping language project. The aim of this project is to ease the development of software by providing tools that allow software structures to be "molded in clay" before they are "cast in concrete."

We aim to put tools and methods in place which will allow developers to produce, at drastically reduced cost, both in money and in work time, system versions that are fully functional and that can be shown to end users (in our environment, to colonels or generals who will have to use the software in the field), enabling them to see and approve the full system functionally, although not on the small computer on which that function may ultimately have to be run.

That is, our prototyping technology will aim to use more machine muscle during development in order to speed up development. We believe that once prototype development is complete, systems can be redeveloped in their final, efficient forms with much greater speed and assurance than would otherwise be possible, and that the whole development process will thereby be improved.

Let me mention one last piece of work, just beginning, which links our software effort to the national network concept that also stands very much at the center of Senate Bill S. 1067.

This last project aims to create a so-called "national file system." The interesting technical possibility here is the following: The gigabit nets that are being put in place will supply about as much data communication bandwidth between New York and Los Angeles as one now has internal to a single computer.

That means that in principle one ought to be able to use, from New York, files that are located, let us say, in Los Angeles or Chicago, and not be able to tell whether those files are resident on your computer system or 2,500 miles away.

We want to put together a software structure that realizes that technical possibility by enabling people anywhere in the country to attach to files anywhere else in the country, thereby creating a seamless national file system.

This points cut one of the many important economic possibilities implicit in the national computer net concept: the notion of a "Library of Congress in every home." The national file system we will be developing will give everybody access to all the books in the world if only these are put in electronic form and made available over such a net.

Senator GORE. Is there objection? We will just do that, then. We will make that a part of the bill.

Go ahead. Excuse me.

Dr. SCHWARTZ. Let me conclude with just a few remarks about DARPA's special capabilities and the role that we see for ourselves within the broad national software technology thrust laid out in S. 1067. In order to work well, any software tool must reflect the innermost structure of the area to which it will be applied.

This is a very basic point; let me try to enlarge on it. The reason that a spreadsheet program like LOTUS 1-2-3, to take a particular example, is so effective is that it captures the accountant's manner of thinking.

Similarly, a programming language for some area of engineering or biology would have to model the mental processes of the engineer or biologist.

In order to understand any such area—and it is this understanding that, as Bill Wulf has said, that is fundamental—one has to think deeply and carefully about the area, forming just the right mental model of what an accountant does, or of what an engineer or biologist does when setting up the kind of computation for which the programming language will be used.

One has to find just the right model; that is not easy. It takes time, it takes thinking, and it takes false starts, which is to say it takes research.

However, once this research culminates in a pretty good model of what one wants to do, software engineering of the conceptual solution must begin, and one must scale up from a relatively small number of researchers sitting around in their offices and laboratories to concentrated programs of software building, which may involve groups of 30 or more people working in a concentrated, coordinated way for several years.

It is in organizing and managing such efforts that I believe DARPA has something quite special to offer, because I believe DARPA has excelled in the management of these larger projects. That is the special role we see for ourselves and I wanted to point that out.

That is all I have to say. I will be happy to take any questions.

[The statement follows:]

STATEMENT OF DR. JACOB SCHWARTZ, DIRECTOR, INFORMATION SCIENCE AND TECHNOLOGY OFFICE, DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

Good afternoon, Mr. Chairman, ladies and gentlemen:

My name is Jack Schwartz, and I am Director of DARPA's Information Science and Technology Office or 'ISTO.' This is the part of DARPA most directly concerned with information technology in all its aspects. ISTO sponsors roughly 100 million dollars per year of computer science research and development work, approximately half in universities, the other half in industry and at non-university research laboratories. The projects sponsored span the whole of computer science and its applications, from design of integrated circuits and of new computers, through development of software tools, new methods in artificial intelligence and robotics, and new computer algorithms, to military and non-military applications of all these technologies. I am also Director of DARPA's Strategic Computing Program, whose activities, budgeted in Fiscal Year 1989 at roughly $125 million, builds on ISTO's program of basic research extensively, but which focuses on transition of the more mature of these technologies to practice and includes projects managed by other DARPA offices. I am most pleased to have the opportunity to testify to you today concerning the general importance and problem of software, on the programs which DARPA has put and is putting in place to address these problems, and on the role which DARPA believes it can most appropriately play in the national software effort that we believe is needed. I note also that software is a problem about which I have been thinking and writing for twenty years, in an earlier existence as a university Professor and industrial consultant.

The physical capabilities of computers, their raw speed and memory capacity, have advanced steadily and with enormous rapidity over this entire period. Software technology, although it has advanced significantly over this period, has not moved forward nearly as much, and this relative (but by no means absolute) failure has progressively turned software into the great unswallowable lump of the entire computer field. This is so for a basic technical reason: It always falls to the software developer to solve any problem not solved elsewhere, in any given system. The 'software problem' is therefore not a single problem: rather, it is the concentrated residue of all the complications inherent in every problem which the speedup of computers and their ever more extensive use has encouraged us to tackle. The software problem should for this reason be compared, not to the problem of reaching the moon, for example, but rather to the convoluted problems which physicians face in prolonging life and health: though people do live longer and are healthier than was the case a generation ago, everybody understands that even if smallpox and cancer are conquered as killers, this will simply bring other causes of mortality into first place, so that the technical struggle must go on. Since software technology struggles to control complexity itself, and rapidly mounting complexity at that, its perspective is just the same. And, from the industrial and defense points of view, software is just as important as strength and health is to the individual because all the management, analysis, or control problems not solved elsewhere always turn up as software problems.

Let me give you a few budget and other facts to put the Defense importance of software into perspective. 'Software producibility' is called out as one of 22 defense-critical technologies in the recent critical technologies plan prepared for the Senate by a DoD-wide working group chaired by Dr. George Millburn. Total national estimated software costs total roughly $110 billion according to Barry Boehm of TRW, both concentrated development and ongoing maintenance costs being included; this represents a national programming workforce of over one million. Annual software expenditures, summed across all DoD programs, are currently projected by the Electronic Industries Association at roughly $17 billion/yr, which is to say that 170,000 programmers work, essentially full time, developing, testing, repairing, and maintaining defense-related software. Closer subestimates reveal more of the inherent messiness of the software problem. According to Boehm, maintenance and rework for elimination of problems or for functional upgrade account for roughly 70% of DoD software costs. Code design and development accounts for only one third of these costs on average, about two thirds being incurred in the preparation of documents. So your mental model of software should resemble the sprawling system of subways, water tunnels, sewers, telephone and power tunnels, and so forth underlying New York City; and the million-strong community of programmers should be thought of as being involved, not only in the ongoing addition of new parts to this structure, but even more in its repair, its extension, and in the unending updating and reading of maps to the labyrinth, so that segments needing rework or repair can at least be located.

Although the advance of physical computer technology has helped the software developer in many important ways, it has also complicated his life by creating ever more complex systems which need to be managed. It has helped, for example, by allowing more and more computing power to be used to process and analyze software systems, and by giving the programmer a personal high-performance workstation environment rich in graphic and other aids that can be used to sift through the very complex structures with which he or she must deal. It has also helped by steadily reducing the need to complicate software designs in order to squeeze performance out of machines barely sufficient to the tasks being attempted. On the other hand, the programmer is now required to deal with new, highly parallel systems in which hundreds of thousands of activity streams must be closely coordinated for full system potential to be realized; must manage worldwide data communication networks that must function as reliably as the telephone system; he must ensure that cash withdrawals made at teller machines in Hawaii are, with perfect certainty, properly debited to parent accounts at banks in New York, Chicago, or San Francisco; and must guarantee that malicious (and possibly very expert) programmers able to access U.S. computer systems from Frankfurt or Singapore can never locate any system chink that allows them to disrupt service, or to steal service, cash, or closely held information.

Let me say something about the programs which DARPA has put in place to deal with these challenges. Between its basic research program and the part of the Strategic Computing program which it manages, ISTO spends about $25 million annually in software research and tool development, including specifically network-related software. Other important software programs, generally having either a shorter term or a technology transfer focus, are in DARPA's Defense Manufacturing Office. These are the STARS (Software Technology for Adaptive, Reliable Systems) program, currently budgeted at roughly $12M, and the major Software Engineering Institute operated for DARPA by Carnegie-Mellon University. The STARS program is developing a comprehensive set of Ada-based tools for constructing and managing very large software systems (for example, those requiring over 500 person years in development); projects of this size and greater must frequently be undertaken in connection with major Defense developments. STARS aims to achieve dramatic improvements in the affordability and quality of these large software products. The SEI continually reviews software technology developments and makes them directly useful to Defense organizations, through a very extensive program of publications, packaged training courses, seminars, and workshops, and specialized consulting services. Although only four years old, the SEI is already recognized for its leadership in establishing a software engineering process improvement program within a large segment of the defense industry, for its efforts in real-time distributed systems technology, and for its expertise in use of Ada. The Software Engineering Institute is currently funded at about $19M annually.

Even though there is no time to say anything in detail about the dozens of individual software projects that DARPA sponsors, I would like to bring a few particularly salient examples of what ISTO has done and is planning to do to your attention. Over the last few years, we have sponsored the development of a very promising new operating system, called Mach, at Carnegie-Mellon University. As you know, these are the software packages which establish the basic environment of screen, keyboard, communication, printing, and file services that any user expects to find on a computer. Among the many advantages of Mach I note the fact that it was designed from the start to be suitable both for the high-performance single user workstations already on the market, and for the large new parallel machines that have been developed in the Strategic Computing Program and its commercial offshoots. Mach was also designed to be 'portable', in the sense that it is carefully insulated from machine dependent details. so that it can be adapted very easily and rapidly to run on new machines, allowing all its users to work in a single, familiar environment requiring little or no relearning even if they switch from machine to machine. It has already been made available on over a dozen different computers, of which two of the most interesting are the NeXT workstation that Steve Jobs has developed and the popular parallel computer developed by Encore Corporation.

ISTO is also sponsoring important work on parallel software tools at Rice University in Houston. This draws on earlier University of Illinois research on methods for efficient generation of parallel programs. Beyond this, we are about to launch a major new activity aimed very directly at the general simplification of software design. This is ISTO's new 'Common Prototyping Language' project. The aim of this project is to make software development easier and much less risky than is now the case by providing tools that allow rapid construction of system prototypes that do exactly the same things that a final software system will do, though not necessarily with the same efficiency. This 'prototyping' approach makes lavish use of computer power, which is an increasingly cheap resource, and substitute it for human labor, that is, the detailed efficiency-oriented design that would otherwise be required during the initial design and development of a software system. Once such a software prototype is complete, and has been approved both technically and by its eventual user, the remaining problem of rework for efficiency can be approached systematically, with critical advantages for manageability and predictability.

The last project that I will mention is work just beginning, which links ISTO's software research plans together with the national very high performance network, described in Senate Bill S.1067. This is a project to create a 'national file system', that is to use the ultra high speed network capability to create a software environment in which computer files anywhere in the national can be accessed by an authorized user with the same ease that you can access a file resident on your own desktop machine. Fully realized, this can create a 'Library of Congress' in every home, as just one of many very exciting commercial and military spinoffs.

I and my colleagues at DARPA agree that software is an area of computer technology whose abiding, indeed inevitably growing, importance will continue to justify major national technical efforts. ISTO has been a very active participant in developing a plan for the Federal High Performance Computing Program and we strongly believe that in order to capitalize on the planned hardware investments it will be essential to continue work on parallel software concepts and tools, and on parallel algorithms. We believe that parallel computing technology, much of which has come out of the DARPA Strategic Computing Program, has opened a major new chapter in the history of computing. Work done to date has given the U.S. a substantial world lead in this area, but there is no doubt that our international competitors are on the move, and that if we are to stay ahead in this race we need to follow through.

Let me conclude with a few remarks concerning DARPA's special capabilities and the appropriate role in a major national software technology thrust which these capabilities justify. To be effective, any software tool must reflect the innermost structure of the area to which it will be applied. For example, a spreadsheet program like LOTUS 1-2-3 is very effective because it captures the style of ordinary bookkeeping calculation so well, and makes this directly available to its user through a well-designed interface. To do this well with other target applications, one has to think carefully about the target application and find just the right approach to it; finding the best rather than the second best approach has sometimes taken years. But, once such a process of conceptualization reaches its goal, one must really build the imagined software tool. Because the inner details of this tool may be complex, doing so typically implies one or more substantial software build projects, which may, for example, require several years of active university/industry collaboration. DARPA's track record shows it to excel in the management of projects of this kind. Our aim is accordingly to cooperate with other agencies during the ongoing research process which must precede any major undertaking, but then to step forward in a particularly active way when the time to build a major software artifact is at hand.

I hope that these remarks are useful to you and I will be happy to answer any questions.

Senator GORE. Well, great. Let me move to some questions then. And again, thank you both for getting us off to a good start here.

The 1987 OSTP report on a research and development strategy for high-performance computing highlights several grand challenges in areas like aerodynamics, global climate change, and computer vision. Programs would be developed within the Federal agencies to develop the software and computing capability needed for researchers to meet these so-called grand challenges.

Do the two of you support the idea of focusing on particular grand challenges?

Dr. SCHWARTZ. I am sure that Bill Wulf does. I would be surprised to hear him say no.

For my part I certainly believe these projects to be significant. Let me comment to explain why. First of all, these so-called grand challenges are scientifically very rich, and ultimately economically very significant, if you look behind the first wave of science.

I would note one other point: when one develops a new machine, necessarily carrying new software, one needs to learn how to use it effectively. To do this it is appropriate to start with an application that is very significant, but at the same time not over intricate from the technical point of view, that is with a problem that is scientifically mature.

The grand challenges have been very well selected from that point of view, so they are ideal things to practice on.

Senator GORE. Which ones do you find most interesting and most important?

Dr. SCHWARTZ. Well, for example, some of the work proposed in chemistry and in molecular biology. I think for example, that any improved degree of understanding of biomolecules is going to have all kinds of consequences both for human health and for U.S. competitiveness in the pharmaceutical area.

Dr. WULF. I would not disagree with Jack. I cannot help but observe that disciplinary scientists from different disciplines would probably choose different grand challenge problems as the ones that are most important.

But all of them have the characteristic that they push forward not only their own discipline, but the discipline of computational science in general, and consequently have a spillover effect on all of the scientific enterprise.

Senator GORE. I would just add that I think the global climate change challenge is one that ought to be right at the top of the list.

Dr. WULF. It is more of—it is not exclusively a computational problem.

Senator GORE. Right.

Dr. WULF. It is a problem which is distributed in space. That is, the phenomena happen at disparate points. It is a problem where the appropriate researchers are distributed.

So it is a problem where the instrumentation to make the appropriate measurements has to be distributed, and often unmanned. It relies heavily on information processing technology, not just the computational technology. It requires remote sensors, networking.

Senator GORE. Right. But we are right now in the midst of a tremendous national effort to increase by many, many orders of magnitude the amounts of data collected in the earth sciences as part of the Mission to Planet Earth program and related activities.

Those designing that effort have long since identified data management as the most difficult aspect of the challenge. And simultaneously, climatologists working with the data now available tell us that better models are needed.

Much can be predicted, to varying degrees of certainty, with our present models. But everyone recognizes the need for dramatic improvement there. So I would just put in my bid there.

Will visualization play a big role in meeting these grand challenges?

Dr. WULF. I think that you made an eloquent case for that point in your opening statement, Senator. The amount of data that results from the kinds of scientific computations that we are talking about is enormous. It is not the sort of thing that human beings can understand by looking at a pile of paper ten feet high.

No intuition comes from that. A famous computer scientist many years ago observed that the purpose of computation is not numbers, it is insight. And scientific visualization is one of the ways in which we can present the information to the disciplinary scientists in a way that leads to intuition, to insight, to understanding.

Senator GORE. If you both agree with the grand challenges approach, how are they going to be funded? Through the computer science divisions of the agencies involved or through other divisions?

Dr. SCHWARTZ. Since most of this funding would fall to NSF and to DOE, I think I will ask Bill to answer that question.

Dr. WULF. In the particular case of the National Science Foundation, some of both happens. We have, as you know, within my directorate the Division of Advanced Scientific Computing, which is responsible for the five national computer centers.

We have just gone through a renewal process for those five centers, part of which strongly emphasized the need for a significant intellectual component at the five centers, not merely a service and educational component, which was their primary function in their first five years.

In addition, in that division we have established a so-called New Technologies program. The title is perhaps not terribly descriptive, but it is focused on computational science and on the grand challenge problems.

One of the things that I have learned since coming to the Foundation, and perhaps did not understand very well before, was that there is an enormous degree of collaboration between program directors in different programs, indeed within different divisions and within different directorates of the Foundation. A great deal of cross-funding goes on.

Although there are specific programs in computational mathematics, in computational chemistry, in computational physics, in computational biology, and so on, we cross-fund many of the proposals that come into those directorates and they cross-fund many of the proposals that come into our directorate.

Senator GORE. Let me come back to visualization for a minute. We are going to be dealing with it a little bit more in the hearing later.

But the model I have in my mind to represent this problem tells me that we as human beings process and understand information in a particular way, and if you attempted to describe the human brain in computer terms you would say perhaps that we have a low bit rate but very high resolution.

In other words, if we attempt to absorb and integrate new information one bit at a time, as a sequence of numbers or a sequence of facts or a sequence of abstract symbols of any sort, the absorption rate is relatively slow. But if that data can be presented in visual form so as to create a mosaic of data, then we can recognize patterns and absorb vast quantities of data much more quickly and much more easily.

I used the figure in my opening statement that half the human brain is devoted to processing images. Some people say two-thirds.

Therefore, if that is correct then it makes sense for us to accelerate the pace of work on software that facilitates the presentation of massive quantities of data in visual form. Let me pause there. Is that correct in terms of your knowledge? Have I gone wrong so far?

Dr. SCHWARTZ. Let me comment. I would make the same point in the following terms: the eye routinely performs miracles greater than any achieved by the rest of the mind. We struggle to appreciate with the rest of our mind the things that the eye picks up instantly.

Vision is undoubtedly one of the most miraculous capabilities that humans have, and for that reason it is entirely appropriate to make every possible use of vision as an aid to understanding.

However, I would add to this the footnote that vision will work best in the area of scientific computation, i.e. of computational science, because it applies most naturally to objects and situations that are naturally part of the physical universe. Example might be the flow of fluid around an aircraft or the shape of a molecule, that is anything that is intrinsically three-dimensional and physical, rather than being intrinsically abstract.

In contrast, abstract mathematical relations are sometimes best understood, not by looking, but sometimes by shutting one's eyes and thinking. So there is the other side to which visualization techniques are harder or impossible to apply.

But vision does carry us very far, and we certainly want to make all the use of it that we can.

Senator GORE. Well, suppose you shut your eyes and thought and then you conceived of the solution to the particular problem you were pondering. If you then wished to communicate what you had conceived you would not tell your listener to shut his eyes and listen to you describe it in verbal terms.

You might, but if you could present it as a picture you would do so.

Dr. SCHWARTZ. If I could, I certainly would. But in some cases I might have to admit that I cannot, so we have to use the other verbal rather than visual approach.

Senator GORE. Right.

Dr. SCHWARTZ. To explain, for example, the notion of "goodness" or of "legality", to take just two abstract notions, I might not want to draw pictures, but instead to rely on words.

Senator GORE. All right.

Dr. Wulf?

Dr. WULF. I thought that you characterized the situation extremely well. I guess the only thing that I would add here is that this is a special case of a situation that we need to deal with in general, and that is that the computational science community is not well supported with tools for constructing their scientific programs.

They need software tools to help them construct visualizations. They should not have to work at a very detailed level of how you get that picture displayed. That is a problem for computer scientists or for computer engineers.

Senator GORE. Right.

Dr. WULF. And that same statement is equally valid with respect to constructing the programs which solve partial differential equations. We need to give them much better support than they currently have for all of these kinds of activities.

Senator GORE. What percentage of your advanced computing budgets do you devote to developing and using visualization technology?

Dr. WULF. I would have to get back. I do not know the answer to that off the top of my head.

Dr. SCHWARTZ. Let me comment. ISTO does not have items specifically budgeted to visualization per se, and perhaps we should—I think you raise a very interesting point—except in one regard, which belongs to what might be called military visualization. Here we do have a large program, which may have come to your attention. This is the so-called SIMNET program, which is a growing network of interlinked simulators that visualize the battlefield.

Senator GORE. Right.

Dr. SCHWARTZ. With all the benefits that that implies.

Senator GORE. Well, we will hold the record for a further response to that question, if the two of you could examine your budgets and try to extract an estimate of what percentage is being devoted to visualization and whether or not there has been a trend over the last two, three years.

[The following information was subsequently received for the record:]

We estimate that approximately 10 percent of the yearly NSF program budget in Advanced Scientific Computing is devoted to the development of visualization technology.

In addition, much of the fundamental research in other NSF/CISE programs is indirectly supportive of visualization—for example in the development of new architectures for computer graphics, for tools used to design, prototype, test, and program such architectures, etc. These are not "single use" research efforts, however, and it is impossible to estimate the fraction that is ultimately used specifically for scientific visualization. It is clear, however, that contemporary visualization would be impossible if it were not for similar basic research in the past.

Scientific visualization is widely used by disciplinary scientists and engineers who perform research at the NSF Supercomputer Centers. In addition, researchers supported by the Foundation that are using one of the state, university or regional supercomputers, or high powered workstations in their own laboratories, are exploiting visualization. We do not have a way to estimate the fraction of our support used in this way.

Senator GORE. Is it possible that the kind of supercomputer graphics that we will see a little bit later in the proceeding might be commonplace in PC's in the years ahead?

Dr. **Wulf**. Absolutely. One of the things that we certainly ought to have learned about the computer and information processing business is that it moves extremely quickly.

I think I have used the analogy before that, if the automobile industry had made the same progress as the electronics industry since 1957, an automobile would cost three cents and, would cruise at—I have forgotten what it is—20,000 miles an hour, something like that.

I think we can confidently predict that the kinds of—oh, and it would be the size of a cigarette package, if that helps you.

Senator **Gore**. The car would be?

Dr. **Wulf**. The car would be, yes. I do not know how you would get in it.

Senator **Gore**. Like a skateboard.

Dr. **Wulf**. But the thing that we can confidently predict is that the kind of visualization that you are going to see on a graphic supercomputer today will be available at PC prices certainly within ten years.

Senator **Gore**. All right. You all have discussed the software bottleneck. It seems to me that this is an area where government policy should focus. If you look at the hierarchy of development, the hardware comes first and fastest; then lagging behind at a slower pace is the software; and slowest of all, lagging far, far behind, is government policy.

We need to identify particular software projects which, as you say, Dr. Schwartz, will have a very high scientific payoff. Then by successfully developing particular software projects we can get past bottlenecks that presently impede fast progress in areas of scientific inquiry and economic competitiveness. I think it makes sense to focus government efforts in this particular area on those priority software bottlenecks that impede progress.

It is my view that the Federal government has been slow to recognize that that is the case and slow to recognize the need for increased funding in advanced computing. You all both agree with that, do you not?

Dr. **Wulf**. Yes.

Senator **Gore**. He is speaking for you, is he not, Dr. Schwartz?

Dr. **Schwartz**. I would add a footnote to that, just to flesh out what I think needs doing. One needs to maintain a reasonably large and healthy group of thinkers who are looking at a variety of application areas and trying to understand them profoundly, in order to find the right software approaches in those application areas.

These people are researchers in a basic sense. They must think about the right way to do things. One wants to make sure that they are encouraged to do so and that they have an adequate level of resource.

One also needs to stay alert for the opportunities to escalate the more successful of those efforts into real software build projects. I think it is more on the latter than on the former side that the U.S. has fallen down in the past.

You know, the NSF and DARPA and the other agencies involved have supported a wide variety of smaller projects, not everything that is needed, but much nevertheless. However, when it has come to really building large packages, there just has not been adequate government support in the past.

Senator GORE. And we really should change that, should we not?

Dr. SCHWARTZ. We really should, yes.

Senator GORE. Well, S. 1067, is designed to help do that.

Excuse me, Dr. Wulf. Go ahead.

Dr. WULF. I just want to say that there are also government policies which can, rather than not fostering, actually impede our progress on software development. One that was pointed out by the Defense Science Board report on military software was procurement policy, particularly the rights in data clauses in current procurement policy. Basically, the current situation is one in which government contracts for software require not only that software be delivered, but also any tools and technology that were used to construct the software.

Now, superficially that is a fairly rational thing to do. It implies that, should it be necessary for it to do so, the government can take over maintenance of the software because it has the tools necessary to do it.

Unfortunately, there is an unintended side effect of that policy, which I believe is disastrous, namely that it effectively discourages the software contracting community from investing in such tools, because the moment that it makes that investment and uses them on a government contract it loses its proprietary interest in that software. It has effectively given it away.

The effect of that I think is profound. Either contractors do not make the investment and so continue to use labor-intensive, nonmodern, old-fashioned techniques or, if they have made the investment, they do not use their best tools and technology on government-procured software and consequently the government gets less high quality software, it pays more for it, and it takes longer to develop than it should.

Senator GORE. Well, I think that is a very, very important point. Just to restate it briefly, what you are saying is that, if a company sells software to the government, then the software used in creating the software becomes the property——

Dr. WULF. Of the government.

Senator GORE [continuing]. Of the government. And so this discourages software companies from supplying what the government needs and benefiting from the challenging work that the government needs to have done.

I might just say that the legislation, S. 1067, which I have introduced and which Senator Pressler has co-sponsored, provides that Federal agencies procurement regulations will be changed so that software companies providing software to the government need not forfeit the proprietary software tools that they used to develop the software.

That is section 401, Title IV, section 401.

But I am glad you brought that up, because we encountered that problem in drafting the legislation. And I am glad you agree that that is one of the things we can do to really speed things along.

I just have a couple more questions and then I want to recognize our ranking member.

Both of you talked about the network. Both of you are strong supporters of the network, is that correct?

Dr. WULF. Yes.

Dr. SCHWARTZ. Yes.

Senator GORE. Okay. I believe that there is nothing we could do that would be more beneficial to the software prowess of the United States of America, because increasing access of one team of researchers to another and allowing the easy distribution of software would really supercharge the various research and development efforts throughout the United States.

I really think that is yet another reason why the national network is so important.

I am pleased to recognize the ranking Republican member of the subcommittee, Senator Pressler.

OPENING STATEMENT BY SENATOR PRESSLER

Senator PRESSLER. Thank you very much, Mr. Chairman.

Supercomputers are our most powerful tool for scientific and technical advances. One of the ways we can use that tool—and I look forward to the demonstrations we are going to see this afternoon—is through software that generates pictures of the data we receive, so we can visualize the data.

Visualization permits scientists and researchers to transform numerical data into pictures on a computer screen. Since half of the human brain is used for vision, it is much easier to understand, interpret, and analyze pictures than thousands of pages of numerical data. This is particularly useful for processing the data we receive from our earth satellites. Our satellites are currently sending us tremendous amounts of data about the environment, geography, geology, weather, and our solar system that need to be processed. Supercomputers and advanced computer software will permit us to process that data quickly and transform it into an easily useable format.

The EROS System near Sioux Falls, South Dakota archives data and pictures from LANDSAT. It soon will also be receiving pictures from NASA satellites on Mars and from foreign satellites. We already receive so much information that it is impossible to process it all. Almost 90 percent of the satellite data that we get goes unused. And as advancements in satellite technology continue, we will be receiving even more and better data. Scientists and researchers studying the environment, agriculture, the ozone layer, and other important topics need that data. The supercomputer network will permit them to be able to use it by transforming the data into easily understandable pictures, and transmitting the pictures to researchers around the country.

Mr. Chairman, to move things along, I shall ask a couple of questions and may have some more questions for the record.

I might say that it is my feeling that a computer network across the U.S. has great potential for making scientific advances in the study of the greenhouse effect and changes in global weather patterns, as well as assisting American farmers and others.

Now, in what way could a computer network—how could it expand our knowledge of the greenhouse effect and changes in global weather patterns?

Dr. **WULF**. Well, Senator, as I commented before, the entire global change problem is an intrinsically distributed problem. Obviously the phenomena are distributed. The instruments necessary to measure those phenomena are necessarily distributed.

The researchers are distributed, and it is necessary for them to communicate with each other, to access the data collected by each other, to be able to execute the models—the computer models that are executed on supercomputers—that each other have created, and to execute them on the high-performing machines such as those at the national supercomputer centers.

So intrinsically this group, perhaps more than any other collection of researchers in the world right now, depend upon communication with each other, data communication, personal communication, collaboration.

Senator **PRESSLER**. Well, I have mentioned before my interest in having the pictures from outer space made available to university students, to everybody, which I think is directly related to the issue of the warming of the Earth.

Let me ask——

Senator **GORE**. Could I interject just a brief point there to elaborate on his last point?

Much of the research on global change is being done by teams of climate researchers, many of whom use the kind of data that is stored at the data center in Sioux Falls, South Dakota. Today, when one team wants to communicate with another team and compare the latest improvements they have made in their models of the climate system, they have to download their model onto magnetic tapes or some other storage medium that is transportable and take it, along with their suitcases, get on an airplane and fly to their colleagues' lab. Then, they have to upload their software on their colleagues' hardware so they can all look at their results and talk about them.

Then they download everything and go back home again. That is basically what they are required to do now?

Dr. **WULF**. Yes, that is right, that is the situation.

Senator **GORE**. So as a practical matter they do not do that very much. If they could instantly share data and cross-fertilize each other's models, they could speed up the process of understanding the Earth's climate system exponentially.

Dr. **WULF**. Notice that this implies not just communication of data, but human interaction across the net, collaboration. I have talked about something I have called a "collaboratory," an intentional pun upon the words "collaboration" and "laboratory", but an electronic analog of the physical scientist's laboratory, in which people who are physically remote from each other can collaborate, share data, share instrumentation, share programs and so on.

That concept, it seems to me, is absolutely crucial for the climatic changes that we are trying to study these days.

Senator **GORE**. Thank you.

Excuse me.

Senator PRESSLER. That is very useful because, as I understand it, now there is a backlog of data that we have from our satellites that has not yet been analyzed, and the advances in computer software would help to eliminate this backlog of data as I understand it, even without getting new pictures down.

A lot of times, if we could make all of this data available to people, that would really speed up research and knowledge of, let us say, a big city's problems of pollution, for example.

Dr. WULF. I do not know the exact statistics, but a frighteningly small fraction of the satellite data has ever been looked at by a human being, much less analyzed.

Senator GORE. Less than ten percent.

Senator PRESSLER. Yes, it is just a tiny amount. It is like the library at Ephesus unopened.

Dr. WULF. Right.

Dr. SCHWARTZ. One can now use high-performance computing to pre-filter usual data. If one knows what you are looking for by way of rare events, you can scan through and, say, throw out 90 percent of the cases, which need not be visually examined.

Senator GORE. With massive parallelism?

Dr. SCHWARTZ. With massive parallelism. And then one only need to look at a much smaller fraction of the cases. This is a technique that we are much interested in in the case of aerial photography.

Senator PRESSLER. Now, some people say that we can just leave to the private sector the development of certain types of software applications.

How does the government decide which types of software applications to develop? Do you focus on basic research? What about the argument that, if someone says there are software needs, that we can just leave them to the private sector?

Dr. WULF. Well, there certainly are some kinds of software that can be left to the private sector. But there are two things which come to mind which certainly need government research.

The first is the development of tools and technology for producing software. That is one level removed or one level indirect from the production of software products. You understand that software is used to produce other software. It is a tad incestuous, but that is the way it is.

The second thing is the whole area of scientific software, the software which is used by scientists to analyze data, to model systems like the global weather, where there really is not a market and therefore there is no private sector interest in developing that kind of software.

Perhaps Jack wants to add something.

Dr. SCHWARTZ. Yes. I have always seen the relationship between government activities and industrial activities, in technical areas in general and this in particular, as synergistic, rather than opposed.

By stepping out, by absorbing some of the front-end risk inherent in new developments, the government makes it possible for the private sector to enter more rapidly than they otherwise could. The private sector needs to operate on nearer term returns than the government can be comfortable with.

So there is a natural progression in many of these areas from DARPA, NSF, and DOE programs to private sector activities. This is a relationship that could be traced in hundreds or even thousands of cases.

Senator **Pressler**. Now, how can we ensure the security of data in a nationwide computer network?

Dr. **Schwartz**. That is a massive challenge. The best technique available now, or rather the best family of techniques, for assuring the security of data is systematic encryption.

Encryption technology has come along very dramatically over the last ten years. There now exist well-understood so-called public key encryption systems, which can protect data very effectively.

To go beyond encryption is to enter into a complex though important research area that DARPA is much involved in, but that is still very challenging and has not yet produced decisive results.

Dr. **Wulf**. I would just add to that that it is common in the press—and it was, in my mind, another mistake that was made in the GAO report which just came out—to talk frequently about "security on the network." In a very fundamental sense, the security issue does not lie with the network; it lies with the computers which are attached to the network.

The case of the infamous worm attack of last fall was a failure of the computers attached to the network, the security of those machines, not the network itself. In some bizarre sense, the network worked extremely well. It propagated that information around just like it was designed to do.

The same kinds of attacks could in principle be made over the ordinary voice telephone network. It would not propagate as fast because the lines are slower. But nobody describes the problem as a security problem of AT&T or MCI or Sprint. The focus on the network, I think, tends to mislead people and, I am afraid, would lead to inappropriate kinds of remedies for the problem.

The problem for which we need a remedy—and it is a very important problem; I do not mean to diminish it at all—is the security of the individual systems connected to the network, not the network.

Senator **Pressler**. Finally, and you may want to answer this last one for the record, not off the top of your head, but let us say that we had the space pictures and satellite pictures available on supercomputers, readily available. What kinds of people would use those and for what purposes?

Dr. **Wulf**. That is in fact a good one to get for the record.

Senator **Pressler**. That would be a long list, would it not?

Dr. **Wulf**. Yes, that is right.

[The following information was subsequently received for the record:]

My colleagues from a variety of disciplines agree that satellite pictures contain an enormous untapped reservoir of information. To analyze them and use them in solving important problems in agriculture, weather forecasting or the like remains an ongoing challenge. Much of the work is in its early stages of development. Future uses of this information hinge on the creativity and insight of individual researchers. Many specific applications are still speculative. The critical need to accelerate this process is to make this information accessible to these resarchers—primarily through high-speed networking and high-performance computers—and give them the opportunity to exercise their creativity.

Senator PRESSLER. Well, I would like you, if you can, to think of all the practical applications, how that would expand. Of course, obviously students and so forth, but I am talking about small businessmen, I am talking about farmers, of course obviously environmentalists, I suppose.

Dr. WULF. Well, I can also imagine value-added companies accessing those pictures to extract data for use by consumers.

You mentioned farmers. I do not know that farmers would directly want to access the pictures, but if there was a company which was scanning the pictures to determine the extent of drought or the boundaries of a blight or how a flood is progressing and likely to progress over the future, it is the company that was extracting the data, that was adding value.

Senator PRESSLER. In other words, you might replace the Crop Reporting Service?

Dr. WULF. Exactly.

Senator PRESSLER. You better not say that too loudly.

Dr. SCHWARTZ. I would add to that that, even if farmers did not, the grain dealers in Chicago certainly would.

Senator PRESSLER. Okay. But if you can give us a list of practical applications that we would not normally think of, that, if this information became readily available to everybody, what would happen, that would be very useful for the record.

I know that is too long a story for one answer.

Thank you, Mr. Chairman.

Senator GORE. Well, thank you. I have just a couple more.

First, on software standards. One of the problems the committee has been made aware of is that a lot of software is written to match the requirements of the hardware on which it is going to be run. Consequently, it becomes extremely tedious and almost impossible in many cases to translate that same program for use on a different piece of hardware with a different operating system.

Many software developers and software users have pointed to the need for easy ways to transfer programs from one system to another, so that if you have a solution for one problem it is available to all the users of advanced computing systems.

For example, the Unix operating system is becoming the standard for scientific computing on scientific work stations. However, the problem is still acute for the most advanced computers, and most supercomputers can only use software written for that type of machine.

When ETA, for example, dropped out of the supercomputer business a lot of researchers were left in the lurch because the programs that they were using would not run on any other supercomputers.

On the other hand, computer designers, especially supercomputer designers, tend to resist standards because the standards limit their ability to design faster and more innovative hardware and software by taking unique and novel approaches to solving problems that come up.

So without getting into what can become almost a theological debate, what do you think is the solution to this problem? We are trying to address this problem with a provision in the legislation, and I would like to know, first of all, do you agree with my statement of the problem? And secondly, do you agree with me that the government should develop standards for software it purchases that would encourage the interoperability of software so it could run on different computer systems with different operating systems?

Dr. SCHWARTZ. Absolutely. The problem is inherently difficult. In my belief, government policies made it considerably worse, because we have carried over from the period 30 years ago an emphasis on hardware costs in procurement regulations which is now totally inappropriate.

If we look ahead three years, we can see that it is going to be able to purchase what is essentially half a Cray, or the processing heart of half a Cray, for $500 to $1,000 as a single chip computer. The thought that one ought to save on that part of system expense at the cost of running up software expenses by many millions is really absurd. Yet that is what a lot of current regulation enforces.

Portability, including portability to supercomputers, is a very important issue. ISTO has just started working with Cray toward the port of our Mach operating system, which if it goes forward as successfully as we hope will give them the same operating system that you would find, for example, on a NeXT machine that you use as a personal computer.

Part of what has to be done here concerns the software development situation. The fact that there have not been substantial enough government software development programs has meant that the efforts of all the individual companies focused on their own products; they have naturally tended to encourage a certain degree of anti-standardization.

Senator GORE. Dr. Wulf.

Dr. WULF. Well, I must admit you caught me with your question, because I was about to give you a no answer until you got to the very end of your statement. The thing that changed my mind was when you used the word "encourage."

The point is that there are certain kinds of what I might call "essential" differences between computers, and they induce essential differences in the software. The way that one forms an algorithm for a massively parallel computer is in an essential way, a fundamental way, different from a traditional supercomputer.

There are other inessential differences, minor differences in operating systems, minor differences in versions of Fortran or whatever programming language you are using. Those are the ones which are inessential and inappropriate, and certainly the government should encourage standards to avoid them or to eliminate them if possible.

But let us not go overboard. Let us remember that there are certain kinds of essential differences and we cannot eliminate those.

Senator GORE. Now, on clearinghouses. Researchers in universities and industry often end up re-inventing programs that already exist elsewhere because they do not know what software is available and they do not know how others might have already gotten past a bottleneck that is impeding their own progress.

S. 1067 calls on NSF to establish software clearinghouses where researchers could go to find research software they need. Do you agree with the need for such a clearinghouse, Dr. Wulf?

Dr. WULF. Well, it seems to me that, as I stated in my opening remarks, the really fundamental problem is to provide software for computational science and engineering. There is a small market there.

Consequently, there has not been a rush for the private sector to supply the kinds of software that are needed. I think it is compounded by the fact that the typical computational scientist or engineer has not even been provided with the kinds of modern software development tools that are available to so-called systems programmers.

It does seem to me that the national supercomputer centers are very natural foci for a number of kinds of activities. Certainly it does not have to be exclusively the national centers, but they are a very natural foci. They are a foci for developing the kinds of tools, visualization tools, as well as other kinds of software development tools, for the computational scientists.

They are a natural foci for developing some of the particular kinds of software that are needed by computational scientists—not the development tools, but the software itself.

They are a natural foci as repositories and a clearinghouse, so that software does not have to be written at all, one would hope.

In our renewal process for the national centers, we did try to emphasize the intellectual component, including all of these kinds of activities. There is no big activity in the clearinghouse sort of idea at the moment, but I think it is a good one and one we should pursue.

Senator GORE. If I wanted to make a list of the software bottlenecks which, if relieved, would yield the greatest advances, how would I go about creating such a list?

Dr. SCHWARTZ. That is a very difficult question, Senator. All I can say is assign it as homework to some of the witnesses here and we can do our best. We have all thought——

Senator GORE. Well, the question contains part of its own answer then. This is a bottleneck right here.

Dr. WULF. Part of the difficulty I have in trying to answer that question is that I see software being so critical to so many aspects of the economy that I would not know where to begin to put the priorities.

Do we want to sell automobiles? Then we have to have computer-aided design, computer-aided manufacturing, as well as the on-board software for automobiles.

Do we want to sell airplanes? Do we want to do modern science? We have got to have software.

So I see the problem as so pervasive that I would have difficulty putting a priority on it, and the list would be extremely long.

Senator GORE. Well, the list of budget priorities is extremely long and it is hard to set priorities. But there are ways to determine what is most important, and surely there are some that make up a shorter list of those that have the biggest payoff, as we can presently project the payoff.

So if the two of you would respond for the record to that question, I would certainly appreciate it.

[The following information was subsequently received for the record:]

I must place my highest priority on the software that is used to produce other software. Although there are obvious needs for software development in specific disciplines, the impact of increasing overall software productivity far outweighs that of local improvements.

I place especially high priority on those CASE tools (Computer Aided Software Engineering) targeted to be used by disciplinary computational scientists. The existence of such software would leverage the development of other, more discipline-specific software enormously.

My second high priority area is the development of symbolic-Manipulation systems that can help the disciplinary scientist formulate and manipulate the mathematical representation of a problem before it is "solved" by numeric techniques.

A third high priority is high-quality "generic" packages and subroutines that can be incorporated into specific problem-specific codes. Packages of visualization routines are one example of such a package.

Senator GORE. Let me say at this point that we are going to move to the next panel, and I would like to finish this one by thanking both of you again for being here and helping us get at the subject in the hearing today. Thank you very much.

Dr. WULF. Thank you, Senator.

Dr. SCHWARTZ. Thank you, Senator.

Senator GORE. Let me invite our second panel up: Dr. Karl-Heinz Winkler, Dr. Raj Reddy, Dr. William Poduska, Dr. David Nagel, Dr. James Clark.

While you all are coming up to the witness table, let me just say by way of introducing all of you that: first, Dr. Karl-Heinz Winkler recently left the Los Alamos National Laboratory to become the Deputy Director at the National Center for Supercomputing Applications at the University of Illinois, Champaign-Urbana, one of the NSF supercomputer centers.

Dr. James Clark is CEO of Silicon Graphics of Mountain View, California. Silicon Graphics is a world leader in building graphics supercomputers.

Dr. William Poduska is an electrical engineer and the Chairman and Chief Executive Officer of Stellar Computer, Inc., of Newton, Massachusetts. Stellar makes graphics supercomputers specifically designed for visualization. Prior to founding Stellar Computer, Dr. Poduska founded Prime Computer and then Apollo Computers, one of the leading manufacturers of scientific work stations.

It is very important to note that Dr. Poduska was born and raised in Memphis, Tennessee.

Dr. PODUSKA. Yes, sir.

Senator GORE. Dr. David Nagel is Manager of User Technologies at Apple Computer, Inc., where he oversees research using Apple's Cray supercomputer to help design better products. Even more than most computer companies, Apple recognizes the importance of software since much of the incredible success at Apple is due to its innovative software.

Then our last witness today will be Dr. Raj Reddy, the Director of the Robotics Institute of Carnegie-Mellon University and President of the American Association for Artificial Intelligence. Dr. Reddy, we have had an opportunity to talk several times in the past, and we appreciate you coming back.

Our first witness is Dr. Karl-Heinz Winkler from Chambana, as one of my friends out there calls it. Dr. Winkler, welcome. We are delighted you could come, so please get us off to a good start with this panel.

STATEMENT OF DR. KARL-HEINZ A. WINKLER, DEPUTY DIRECTOR FOR SCIENCE, TECHNOLOGY, AND EDUCATION, NATIONAL CENTER FOR SUPERCOMPUTING APPLICATIONS, UNIVERSITY OF ILLINOIS, CHAMPAIGN, IL

Dr. WINKLER. Thank you. I am honored to be asked to address this hearing, and I followed the discussion so far with fascination.

I think the subject discussed in S. 1067, the National High-Performance Computer Technology Act of 1989, is of supreme importance and central to the future economic development and technological leadership of the United States.

In my testimony I would like to address why that is so and what are the characteristics and issues of a well-balanced national digital computational infrastructure.

What I have concentrated on with my colleagues in the project at Los Alamos for the last five years is to find out ways of how we as scientists and researchers can increase our productivity.

Because computers are such universal tools and have gained so tremendous power, I think they have far-reaching applications for all sectors of society. In particular, we observe an ongoing transition from an analog to a digital-based society. In this context, let us just look at the printing industry, audio, video, television, telecommunications, banking, science and engineering, product development, or even manufacturing and distribution.

The main point to make is that not only do we see a change in the machinery we use—from analog-based machinery to digital—but we are also entering or are in the midst of entering an information age. That is why it is of great concern to me as a working scientist if I want to explore an idea, that first I have to learn what other people have done about the subject already.

As long as the information basically stays in printed form in books and journals in the libraries and I have only access to a catalogue index of it, but not to the information itself, my work is much harder.

So, in this context, to be able to access on-line, for example, the information in the Library of Congress or other libraries, including scientific libraries, would be a tremendous increase in user productivity.

Another transition we are experiencing, as pointed out by the previous speakers, is that computational science emerges as a new and important branch of science complementary to experimental and observational activities and theory. Both of these areas I think are of utmost importance for the productivity of all sectors of society.

What I now would like to do is approach the subject from the point of view of a practicing computational physicist and see where there are other roadblocks in the system. This session is in particular dedicated to visualization, and as we have heard, our human brain-eye system is one of the miracles of nature.

In fact, it took nature about 500 million years to develop it. Vision was first developed in early fish about 300 to 500 million years ago.

What makes it so special is that through the visual process if you observe our language, you visualize something and you speak figuratively. In our culture, it is really ingrained to deal with visual imagery.

Senator GORE. I see what you mean.

Dr. WINKLER. We, I think, can best in many areas—there are some exceptions—couple machinery and information to the human thought process. Humans have by far the best pattern recognition activities compared to any system we ever built. We have the ability for abstraction, and in this context our ability for abstraction is a data compression algorithm.

As I will show you later in the video tapes, I can take a complex simulation which took many hours of supercomputer time to produce and then, regardless of it's production—fast or slow—I would like to be able to interact with that data base on a human time scale. Unless I can do it on a human interactive time scale my mind will wander off and I will think about something else.

So it is very important that we get the interactivity into the game, because that, I think, is the most crucial factor in a productivity increase.

To set up such a national system including the proposed three gigabit per second network as an integral part, requires communication and networking, supercomputer memory and processors, and one thing which I did not see specifically addressed in the bill. This is analog to digital conversion of data, which will be a tremendous enterprise, because one actually would physically have to scan through the information which we do not have in digital form.

Senator GORE. What kind of applications? For example?

Dr. WINKLER. Books, journals, audio, music, videotapes—everything which you have collected in the Library of Congress.

I recently heard an interview of the Librarian, of the Library of Congress and he pointed out there are about 5,000 people employed and 2,000 of them do cataloguing of new items. I would like to have access to those new items as well.

In fact, he pointed out that on some areas of the world, like South America and Africa, you have a more complete selection than those countries themselves. And I think the knowledge hidden in these analog data bases does not get fully tapped.

Society seems to be generating new knowledge at the rate of doubling every 10, 12 years. So that would be a realistic time period over which I think one could convert the existing inventory of analog into digital information.

Senator GORE. I heard someone at the Library of Congress say that each year the Library receives over 5,000 periodicals from the nation of India alone.

Dr. WINKLER. Yes.

Senator GORE. And the total is just so large as to sound unmanageable.

But excuse me for interrupting. Please go ahead.

Dr. WINKLER. So I think tapping really into this knowledge base is very important for all sectors of society, including scientific progress.

Going in hand with this subject is the problem, which I think is pervasive throughout the computing centers I have used so far. High-speed computing machines, large memories, are sort of very glitzy. They are also, of course, extremely important. I mean, we all have something like the grand challenges problems on our minds, where in terms of computational resource and memory we need tremendous improvements.

But one area which is often overlooked is data storage. We need man-machine interfaces, visualization work stations, and we need information generation and extraction software. We need, in order to really build an integrated network and operating system hardware, software, and data format standards.

Let me come back to the storage problem. By reading through the bill, S. 1067, the impression I have is that out of it speaks a vision, which I share, and, I think, a clear and outstanding understanding of a balanced approach to this whole area. The bill attempts to set up a comprehensive system solution, and I cannot overemphasize the benefits of a broad system-wide approach.

What happens when you have an unbalanced environment? Say, you have concentrated on one area, high-speed computing, but you do not have the storage capacity and the speed, or you do not have the visualization, then what you really have done is, you have created a situation where scientists adapt to the system, rather than studying the problems according to their importance.

From this point of view, I do not think that in all the places I have worked in my life there ever was the necessary storage capacity and the access rate so that one actually could take full advantage of the computer cycles available.

Another problem on a larger scale in this society seems to be, as came out in the previous part of the hearing, that there is a tremendous emphasis on data collection, satellites and so on, but in the final analysis there is relatively little effort spent in following up and really digging out the information which is hidden in the data.

Just to give you an example, the ozone hole over the Antarctic was discovered by a British researcher who went there every year. He shot a laser beam into the air and he made measurements.

Then he published, I think, a paper in Nature or in a similar journal and he pointed out that there was the ozone hole. NASA subsequently looked into their satellite data. They had collected data already for six years, and then they made these beautiful movies telling us about it. I think there is a built-in unproductivity in our society which needs correction.

Senator GORE. Information that is collected and stored, but never fires a single neuron in a single human brain. It exists totally outside human awareness. I have tried to use the word "exformation" to describe that particular substance.

There are vast quantities of it, and we are affected by its existence even though we never internalize it. The very fact that it is there means something to us, and we have to find ways to use the advanced computer systems and advanced software to presort, to take crosscuts at it, to look for significance, and extract knowledge that is important to us.

But you are so right that now there is this enormous mismatch between the amount of data or exformation that is collected and stored and the amount of information that we actually look at or think about or use in any way.

Excuse me for interrupting.

Dr. **Winkler**. I think what I should do at this point is to present some examples of the work we have done, to show you the orders of magnitude of performance increase we have to achieve over a normal working environment of a scientist.

I think the bill, S. 1067, by setting up the high-speed network will get us a large step closer. If we look at the human brain-eye system, then we can estimate an upper limit of what is the highest data rate with which we could possibly absorb information.

Senator **Gore**. Say that again?

Dr. **Winkler**. We can look at the human brain-eye system and ask the following question: What is the bottleneck? Are the humans the bottleneck in the data transformation game, or is it the technology, the machinery around us?

What I realized several years ago and published some papers on is that we humans are the limiting factor. The technology has matured to the degree where we can throw data at humans at such a high data rate, that we are limited by our own built-in machinery.

We can estimate the resolution of the eye, its field of view, the depth in terms of colors we can see, in shades of grey and light, and the frequency with which we can distinguish images. Humans can distinguish individual frames, for example, up to a rate of eight to twelve frames per second.

So if we put all these numbers together, then we really arrive at a data rate of a few billion bits a second, i.e. gigabit per second. And so from this point of view——

Senator **Gore**. How many gigabits per second?

Dr. **Winkler**. I would say, depending whether you include stereo vision or not, it is of the order of a few gigabits, certainly less than 15, maybe less than three. And somebody probably can give you an answer that, if you really explored data compression, things like that, that you can go a little bit lower.

But I think it is of order of a few billion bits a second.

But now what is interesting. The data rate proposed in the bill, three gigabits per second, I understand to be the data rate across an individual line. That is what would be required to really take high-resolution data in digital form, put it on the line at one end of the country, and have a researcher, a single researcher, looking at it at the other end of the country.

So it gives you this capability. But in terms of capacity, obviously many, many lines are needed in a high-speed network.

Senator **Gore**. Well, let me just say that three billion bits per second is what the leading network people advised us is an appropriate target. But I have said from the beginning that I certainly anticipate that in the future networks will use new technology to have much larger data flows than that.

In the Mission to Planet Earth program, for example, if I am not mistaken, is expected to produce surges of data at rates of up to 20 gigabaud.

Dr. **Winkler**. Yes.

Senator **Gore**. Excuse me. Go ahead.

Dr. **Winkler**. Probably one point I should make is that when you talked about data rates, I think you spoke in terms of bursts, that is the maximum which can go over this line. When I use data rates, I talk in terms of a sustainable rate.

Senator GORE. I understand, I understand. And the bursts in the Mission to Earth program are different from the sustainable rate.

But what I was referring to earlier is my hope that the sustainable rate would be higher than three gigabits.

Dr. WINKLER. I agree, I agree.

Can we have the first slide, please.

[Slide shown.]

Dr. WINKLER. Now, here we have a series of simulations. What you see are nine different experiments. We have taken a bubble of gas and stuck it in a soap film, suspended it in air, and then ran a shock across it on the computer.

What you see in the left column are the results of a mildly supersonic shock; in the middle when you increase the speed of the shock; and to the right when you increase it even more.

And then if you go to the bottom line, we see the results of a helium bubble, which is lighter than air; in the middle it is fluorocarbon; and up at the top line it is sulphaxofluoride.

What you see is what happens to the gas bubble itself, and also what structures are being formed. The blue areas are vortices, and we discovered there some supersonic vortices. The yellow areas show how the bubble gas gets stretched out.

What I now would like to point out is the following. In the experiments, which were first done in a real laboratory at Cal Tech, they could only do the left column, that is very low Mach numbers. The reason why they couldn't go to higher Mach numbers is that they would begin to melt the plastic parts in the shock tube.

So we extended the area, which is easy to do on a computer, and then we discovered a new phenomenon which was really missed in the experiment. We discovered a supersonic smoke ring, sort of if you smoke a cigar, blow out the smoke, and it actually propagates at a supersonic rate.

By going from one part of parameter space into another one, we discovered a new phenomenon. That is a very important point to make, that you do not perform a single simulation, but you do a whole series of them to study the solution space permitted by the basic laws of physics.

If I may have the videotape, please.

[Videotape was shown.]

Dr. WINKLER. What you see here is a simulation where the little blue ring to the right is the supersonic smoke ring.

What we do first is we study the bubble itself. Here you see density represented by many different colors. To the left in the color bar is low density, to the right high. And you see how the bubble holds up.

We have done a two-dimensional simulation so only the upper half was computed, and the lower half was included to make it easy for your brain to interpret it.

What you see here is a system of shock waves, and you begin to see extremely rich structures, and we will begin to slowly zoom into the new structures we found. Right there the vortex is created. Now we zoom in on it.

There you see another shock now, now coming there, how it comes, how it interacts. It gets intersected.

There is a richness of phenomena, and the data we have gone through so far probably corresponds to, on the order of a gigabit. And so if you wanted to print this out on a stack of paper, it probably would pile up to the height of the room.

Here now, there you see vorticity itself, and there you can see, there suddenly how the shock at the front separates. So what happened was the major difference of a supersonic compared to a subsonic phenomena.

In the next movie we see entropy, which is another variable which we can look at. It shows all the complicated physics very well. There the vortex is generated. There it goes. There it separates, and then it behaves like a subsonic structure.

For example, if you take a satellite image and you look at the Gulf Stream then you actually see vortices being spun off from the Gulf Stream, and they rotate around one way or another, and they maintain their temperature for about a few months. They maintain their identity, so they carry plant life with them. In contrast, if those vortices were supersonic they could not maintain their identity because they would include continuously new water.

What you see next is a series of supersonic jets. We just injected a beam of gas at very high speed. Compared to the previous jet simulation this one shows a factor of 50 accuracy improvement, about five coming through a better method and a factor of ten through more computer time.

When you inject the gas not straight, but you wiggle it a little bit, then suddenly the cocoon extends much more and it propagates much slower. So we had maybe a few percent perturbation and the propagation speed slows down by 30 percent.

What you could see was that the jet was choked, and in the real world, in a three-dimensional simulation we would find that would terminate the length of the beam.

Here now, that's the last movie I would like to show. There we see a jet being shot into a gas which is a factor of 100,000 denser than the gas which we shoot into it. And what is now interesting is every second as you watch this movie here in real time the computer had to do 10,000 time steps.

The computer time for this one simulation alone is about 400 hours of machine processor time. What you see also is the discovery of a huge cocoon which really forms a big cavity at the back end where we shot the gas in.

The beam, the green stuff in the middle, periodically breaks down. The gas comes back. We also have discovered here two different time scales. What I showed you is one minute of videotape in which the bow shock or the beam propagates across the screen. If you wanted to understand all the shock systems in the event, it would have taken an hour, and of course that does not fit here.

So I think I would place tremendous importance. In terms of future visualization software on extending are calculation to higher dimensions, to three dimensions.

By the way, what you saw here was photographed live with a TV camera from a digital screen. So it can be done. If you wanted to do the same type of visualization in three dimensions and time-dependent, so that you have a data frame rate of 30 frames a second and you have high resolution, then you require processing power which is orders of magnitude more than really currently available on supercomputer machinery.

In terms of aiding visualization even more, I think standards will be absolutely essential. If we just make a comparison. We have heard here the example of a car before. Nobody would care to imagine a world where the clutch, the brake and the gas are in different locations when you go from one car to another one.

I think it is at that fundamental a level to really allow and encourage the development of standard graphical user interfaces throughout the various products by the various companies in the industry.

Thank you.

[The statement follows:]

Testimony

Dr. Karl-Heinz A. Winkler
Deputy Director for Science, Technology, and Education
National Center for Supercomputer Applications

Mr. Chairman and members of the Senate subcommittee on Science, Technology, and Space. I am honored to be asked to address you on the issues of advanced computational science. I am Dr. Karl-Heinz A. Winkler, and I am in transition from a staff member at Los Alamos National Laboratory to a Professor in the Departments of Aeronautical and Astronautical Engineering, Mechanical and Industrial Engineering, and Physics at the University of Illinois at Urbana-Champaign. My most recent responsibility at Los Alamos has been Director of the Numerical Laboratory. At the University of Illinois I am also the Deputy Director for Science, Technology, and Education of the National Center for Supercomputing Applications (NCSA), one of the five National Science Foundation supercomputer centers.

In West-Germany I have studied Physics at the University of Goettingen and worked for nine years at the Max-Planck-Institute for Astrophysics near Munich. I first came to this country eleven years ago and have been associated with the Los Alamos National Laboratory, the Lawrence Livermore National Laboratory, and the University of California, Berkeley. I have served on several advisory committees to the National Science Foundation and some of its supercomputer centers as well as the National Institutes of Health. Recently, I have also accepted an invitation to be the visiting scientist at the Numerical Aerodynamics Simulator project at the NASA Ames Laboratory.

Since my immigration to this country about five years ago, I have concentrated on conceptualizing and establishing a constantly evolving, well balanced computational environment for scientific and engineering purposes at the Los Alamos National Laboratory. The main emphasis in our project has been to discover how use of advanced computing technologies can increase an individual researcher's productivity. Our project, which led to the formation of the Numerical Laboratory as a part of the Advanced Computing Initiative at Los Alamos, has convinced me that to increase substantially the productivity of all sectors of society we must establish a **"Digital Computational Culture"** in the United States.

You are now discussing the bill S.1067, **"National High-Performance Computer Technology Act of 1989."** I consider the subject addressed in this bill to be of supreme importance and central to the future economic development and technological leadership of the United States. In my testimony I would like to address the following points:

1.) **Why is the Subject of the Bill S.1067, "National High-Performance Computer Technology Act of 1989" so Important?**

2.) **What are the Characteristics and Issues of a Well Balanced National Digital Computational Infrastructure?**

3.) **Recommendations.**

Ad 1.) Why is the Subject of the Bill S.1067, "National High-Performance Computer Technology Act of 1989" so Important?

The universality and computational power of today's computers has far reaching implications for our society, in particular for

- **the already ongoing transition from an analog to a digital based society,** and
- **the emergence of computational science as a new important branch of science.**

There is no question that we are in the midst of entering a digital era, be it in printing, audio, video and television, telecommunications, banking, science and engineering, product development, or even manufacturing and distribution. All of these areas have in common that they already depend or in the immediate future will depend on data processing in digital form. Where once analog machinery, such as the printing press, etc. was the main tool, we now have to rely on digital machinery, i.e. on computers, to be competitive. Key to our culture is the storage, processing, and exchange of information in digital form.

A major strength of this country has been and continues to be its ability to produce the best and the fastest supercomputers and scientific workstations, the most sophisticated networks, and the most innovative software. Of course, Senator Gore is correct by pointing out that in the same way as the interstate highway system was a key ingredient for the economic development of this country in the last half century, establishment of a **high performance national digital computational infrastructure** will be the key for the future economic development of the United States or, for that matter, any other developed country.

Indeed, it would be ironic, if the United States as the technology leader in these areas, did not take advantage of the digital high technology machinery it produces to build up its own national digital infrastructure, including a **digital high definition television and video system** and not an analog one. The build up of the national infrastructure has to be based on our areas of strength not weakness.

Another way to substantially increase our productivity is the use of **computational science** in scientific and engineering areas to perform basic and applied research, to shorten product development cycles, to improve product quality, and to streamline the whole manufacturing and distribution process. In particular, since its beginning five years ago, the very young NSF supercomputer program has already produced of order ten thousand knowledgeable researchers in universities who now can take advantage of their newly acquired computational skills. These people are a key component of the national manpower pool. The tremendous success of the NSF supercomputer program including its educational component, in part, was initially made possible by the national DOE Laboratories who freely shared their computational know-how and advanced system software. After all, the place where supercomputing was developed into a powerful scientific and engineering tool has been at the national laboratories.

The only option the United States seems to have in the long view to preserve and improve its high standard of living is to maintain its current technological leadership position and increase the productivity and effectiveness in all areas of society. To achieve this we must create and foster an environment and infrastructure from which a digital computational culture will emerge. In this culture, use of computational methods and digital exploitation of our knowledge base will be widespread and accepted as the rule, not the exception. The impact of a well balanced national high performance digital infrastructure will not be limited to computer scientists or to technologists, but will have a positive widespread impact on all people of the United States.

Ad 2.) What are the Characteristics and Issues of a Well Balanced National Digital Computational Infrastructure?

A system approach to a balanced national digital computational infrastructure has to deal with these components:

- communication + networking
- supercomputer memory + processors
- analog to digital data conversion, data collection + storage,
- man-machine interface, workstations
- information generation + extraction software
- integrated network operating system
- hardware, software, and data format standards

At the heart of such a system is, of course, the **National Research and Education Network**. It will connect certainly hundreds of thousands, may be even millions of users. It will give them access to powerful **centralized computing centers** and huge **digital warehouses of information**. Accessing efficiently and productively the knowledge hidden in our libraries, where typically information is stored in analog form on paper, film, videotape, or audio records will require a major effort in **analog to digital data conversion** and an enormous investment in high speed (gigabit/s data rates) and large capacity (terabyte to exabyte) on-line and off-line **digital storage systems**. This last technology will only be fully developed if it is driven by the needs of our entire society. Build up of a **digital high definition video and television system** will achieve exactly this. It will also aid in establishing a **man-machine interface** which allows for an effective coupling of the computational environment to the **human thought process**.

Vision developed first in early fish about five hundred million years ago. Exploiting the human visual system for the transfer of scientific and engineering information allows us to tap into **500 hundred million years of evolution**. Humans have a better abstraction and pattern recognition ability than any man made machine by many orders of magnitude. Computers on the other hand, deal best with binary information, i.e. the numbers zero and one. The challenge then becomes how to translate information from a representation best suited to machines to one best suited for humans. **Visualization** software provides this translation from numbers to images which constitute a common language between observational, experimental, and computational science. To use this common language effectively sequences of images must often be animated. This requires very high network bandwidth. An upper limit of the data rate with which humans can visually receive information is of order gigabit/s. The proposed National Research and Education Network would meet this requirement and establish a **man-machine impedance match**.

The National Research and Education Network has another important role to play, namely that of helping to bring the existing national infrastructure into **balance**. The infrastructure at present is unbalanced for three reasons. First, essential ingredients of an effective system are missing. In particular, a large part of our existing knowledge resides in analog form in libraries, such as in the **Library of Congress** or in the **National Library of Medicine**. Second, even in those areas where information is being generated in the required digital format we do not have the facilities to store it, much less than to make it readily accessible. Third, many of the facilities which already exist cannot function to their potential because they are not fully **integrated**. This neglect denies us a selection of computational research programs based purely on their importance. It puts undue emphasis on those programs which match the system. Thus, a well balanced digital national infrastructure will help to produce a more balanced computational science program.

Only **hardware, software, and data format standards** and a **network operating system** built upon them will permit the creative use of the entire national infrastructure. Creation of this integrated national system by the bill S.1067 can be compared in more than one respect to the Apollo program of the 1960's. The Apollo program was immensely successful in inspiring a whole generation and in driving technological innovation. The success resulted because a program was put into place to develop the technology necessary to do the job. The program was not embarked upon because the technology was available and a trip to the moon was a good use for it.

Within the general context described above I now would like to discuss some specific issues.

Standards: The development of and adherence to standards, both for hardware and software, is of paramount importance in promoting innovation and quickening the pace of news systems integration. Furthermore, the use of standards can remove artificial roadblocks to innovation and progress, allowing full exploitation of an investment in a particular technology and an increase in reliability of new products. The Federal Government can promote the introduction and use of standards, not by assigning the definition of a standard to a governmental agency, but rather by promoting the development of a standard by funding an advanced technology research and development project, and by promoting the use of a standard via requirements in the specifications for procurements or development projects.

An example of a highly successful hardware standardization effort is the **High Speed Channel (HSC)** developed at Los Alamos National Laboratory. The entire development was driven by the needs of a specific application area. Subsequently, HSC has been embraced in the standardization process by many hardware manufacturers, and will hasten the arrival of very high speed interconnections between systems from different vendors.

The Apple Macintosh system software has shown the tremendous advantages that a graphical interface between human and computer has to offer in terms of ease of use and user productivity. Software developers were quick to understand the beauty of the interface in combination with the operating system conventions that allow standard exchange of information between various application packages. But it was this tight set of conventions, coupled with Apple's insistence that software developers adhere to the standards, that made possible the explosion of fantastic productivity software now available for Macintosh computers addressing a wide range of application areas from science and engineering, to business, to home use.

There are separate movements by individual companies to retain ownership and licensing rights to the very basic graphical interface ideas which need to turn into standards applied throughout the industry. We would not care to imagine a world in which, for example, automobile manufacturers had varying specifications for the placement of accelerator, brake and clutch pedals. And yet, this is often the situation software and hardware developers find themselves in today.

Software: An important goal is the availability of on-line, interactive graphics as an analysis tool for computational scientists. Graphics displays must be considered as much more than methods of conveying the results of observations or simulations of physical processes. The use of interactive graphics can have an immediate effect on research in allowing the investigator to interact with a simulation of a physical process. This provides an exceptional degree of insight into both the physical processes involved, and the performance of selected algorithms and methods used to model those processes. High speed network connections and high resolution frame buffers, have made interactive graphical analysis possible.

The government should now not be reticent to fund basic research in visualization techniques and algorithms, as well as seed grants for the development of software tools to quickly capitalize on the investments in the new hardware. The National Center for Supercomputing Applications has had great success at beginning the development of a set of software tools for Macintosh, IBM PC, and Unix workstations. Because of the initial success of these tools and wide acceptance through public domain distribution, continued development and productization has been supported by a number of grants from vendors. This strategy for funding of basic research, prototype development, and productization can be applied to all kinds of new tools to enhance the connection between software vendors and new applications, and to bring new tool products to researchers more quickly. These advances have the secondary benefit of being a focal point around which researchers of various disciplines are brought together to solve common problems.

Parallelism: Current conventional supercomputers are built around the fastest single processors that can be built. But we are reaching an era of diminishing returns for continued development of this type of supercomputer. Indeed, basic limitations of the physics of electronic devices require very large investments in time and dollars to achieve only modest improvements in performance. What we are beginning to see now is the development of computers with many thousands of individual, relatively slow processors and the necessary communication connections to allow transfer of data and instructions between them. These machines can today outperform their larger conventional supercomputer siblings for selected applications. But the potential exists for machine performance increases of several orders of magnitude by improving the speed of individual processors in a massively parallel system and by dramatically increasing the number of individual processors that can be simultaneously applied to a given problem. Such computers will make large gains in cost effectiveness as processor chips become a commodity item, and as efficient software and algorithms are designed for them.

Collaboration, Integration: Fostering close collaborations between computational scientists, computer scientists and vendors in the development and application of new computational resources will be of great benefit. No longer does the computational community have the luxury of accepting a new system from the computer industry first, and then ask what the applications to be run on the machine will be. This approach, which has been observed in the past, can result in mistakes, slow software availability, and even additional hardware development cycles. What is needed is a more efficient coupling between new hardware technology and real applications. This can be realized by careful integration of the efforts of the computational scientist, the computer scientist and the vendor engineering staff throughout the development and early production use of a new machine.

The same concept can be applied to software development. With the rapid availability of supercomputer hardware to the academic community with the establishment of the NSF supercomputer centers in 1985 there was an immediate recognition of a lack of functional applications software and tools. The centers were originally funded to do only operating system and *existing* applications software installs and maintenance. Some centers have raised money to do software prototypes and have established partnerships with software vendors to do further development and productization. There must be continued growth of partnerships of vendors with the national centers and national laboratories to develop, productize and distribute the software required to fully exploit the new hardware and system capabilities.

Ad 3.) Recommendations

Bringing the computational environment and culture to new levels requires a balanced attack since lack of attention to any one area can create a bottleneck which reduces the impact of investments in other areas. The bill S.1067 is outstanding in its understanding of this balanced approach. It attempts to set up a comprehensive system solution. I cannot overemphasize the benefits of a broad, system wide approach. However, in reading the current version of the bill, I find there are some areas which need more attention.

* The bill S.1067 does not address the need to convert present large inventories of information from analog to digital form so that they can be accessed electronically over a national network. I therefore recommend that the appropriate institutions, such as the **Library of Congress** and **other libraries**, be encouraged to embark on the necessary **data conversion** efforts.

I also recommend that for these efforts and for the activities envisioned by the bill **on-line data storage facilities** be specifically funded by the bill. Furthermore, data generated in the course of federally funded scientific research should **be made available via the network as part of the publication process.**

- A **Standard Graphical User Interface (SGUI)** across vendors should be developed which must include **visualization tools**. This will substantially increase user productivity by coupling the human thought process more effectively to our computational environment.

- In my opinion, there is **too much emphasis on artificial intelligence** in the bill S.1067. Artificial intelligence efforts are currently frustrated by insufficient computer performance. We should continue to support research in artificial intelligence, but we should be aware that the other measures described in the bill S.1067 will very likely have a more direct benefit for our society.

- The bill S.1067 must more properly reflect not only a balance of hardware growth and capacity, but must promote a **balance of involvement** of computational scientists, computer scientists, and hardware/software vendor engineers. In addition, it must insist upon a more balanced governmental interagency participation.

Senator GORE. All right. Thank you very much.

Our second witness on this panel is Dr. James Clark of Silicon Graphics in Mountain View, California. Welcome, Dr. Clark.

Thank you. Please proceed.

STATEMENT OF DR. JAMES H. CLARK, CHAIRMAN AND FOUNDER, SILICON GRAPHICS

Dr. CLARK. Thank you, Senator Gore. It is a pleasure and an honor to be here and asked to testify on the merits of this plan, this bill before the Senate.

I would like to spend a few minutes discussing my background because I think that it might help you in your effort to get this bill listened to, I suppose.

Senator GORE. Passed.

Dr. CLARK. Passed.

First of all, about 15 years ago I began to get involved in three-D visualization as a graduate student at the University of Utah, which was a center for that activity. And it is largely with that background that I come here today.

About ten years ago I was a principal investigator on a DARPA contract at Stanford. That contract was roughly $12 million over four years. I was there for three of those four years.

What is interesting about the contract is that what was only $12 million of funding there yielded three leading computer companies in the world today. In the last eight years, three companies have sprung from that specific effort on my contract.

One of those companies is SUN Microsystems; another company is the company I started, Silicon Graphics, which is the leading 3-D visualization company in the marketplace today, with many systems at the supercomputer centers around the country.

But most importantly, those three companies combined revenues today, in just a little under eight years, is $2.5 billion per year, 40 percent of which is to international markets, with a growth rate of about 60 percent per year.

Now, that came from just $12 million in funding. So I am a big proponent of funding of the type that you are putting forth here.

I was asked to comment specifically on the merits of the software portion of this bill, and I will comment on those. I would also like to comment on the merits of the bill as a whole.

The first thing that I have to say is specifically why fiber optics is important. I think this has been covered here fairly well. It is capable of extremely high bandwidths, of a type that simply are not available from standard analog cable technology.

At AT&T Bell Labs there are research efforts on fiber, transducers and repeaters that deliver 200 to 300 gigabits per second. So I think it is clear that fiber is the communications technology of the future.

Fiber is a digital medium rather than an analog medium. We all heard the merits of that.

One of the most astounding things over the last ten years is that we have seen an increase in computer power of approximately a factor of 500 to 600 in price-performance. That is, in the most common types of computers in use today.

And at the very same time, the increase in the most common type of networking performance has been roughly a factor of three. ETHERNET in its original form was a three megabit per second technology, and today it is a ten megabit per second technology.

It is grossly mismatched with the performance capability of computers. So fiber is the key to being able to allow computers to communicate at a bandwidth which is more commensurate with their internal processing capability.

Senator GORE. Now, over what period of time?

Dr. CLARK. Ten years.

Senator GORE. So let me underscore those figures. Over the last ten years the information processing capacity of computers has increased by a factor of 600 times?

Dr. CLARK. Roughly, in Price-performance. Some of that comes in price and some comes in performance.

Senator GORE. Right. For the same price, the capacity has increased by a factor of 600. Over that same period of time, for the networks——

Dr. CLARK. That are most commonly used the price-performance has increased. By about a factor of three to five.

Senator GORE. So 600 times for the computers but only three times for the network.

Of course, my view is that that is partly because the networks are ventures that require coordination and cooperation and initiative by more than one single company or individual. And that is why government needs to play a role.

Excuse me for interrupting. Go ahead.

Dr. CLARK. Absolutely, I agree with that. So computing has outpaced networking dramatically.

I would like to just pose the question, is the three gigabit per second rate proposed in S. 1067 sufficient? I understand the comments you made earlier, but I want to just point out that, a gigabit is roughly the amount of information on an audio CD; it is roughly one second's worth of high definition television.

Today, 40 personal computers have an internal capability of manipulating data at a gigabit per second. In addition, two modern RISC work stations that cost roughly $10,000 are capable of manipulating data internally at this rate. And machines of the type that both my company and Mr. Poduska's company sell are capable of saturating a three gigabit network, just two machines.

So I think the three gigabit target needs to be thought out carefully in terms of the longer term implication of that. I have heard it said that you refer to this as the interstate data highway system. I think that is a very apt description of what is needed. What worries me is that three gigabits could end up being basically a country back road in 1996.

Senator GORE. Well, I fully agree, but it is an intermediate goal.

Dr. CLARK. It is an achievable goal.

Senator GORE. It is an achievable goal.

Dr. CLARK. Right.

Senator GORE. And once the network is in place, the fiber itself has capacities inherently far greater than three gigabit. Your limiting factor is in the electronics and the software. As you get users on the three gigabit network, then you get the potential for pay as you go financing for continual upgrades to larger capacities as the nation expresses the need for those larger capacities.

I agree with you that three gigabits, although a huge improvement over the present capacity, is really only the first step toward data flows that will be much higher than that.

Dr. CLARK. I agree. Part of my comments have to do with ways in which this country might be able to leverage consumer uses of fiber optic networks.

So in that regard I would like to move to the next topic which I have to discuss, and that is high definition television. As I mentioned earlier, a high definition television signal of the highest possible quality is roughly a gigabit per second.

As you know, there is substantial effort under way in industry and a fair amount of confusion about what to do in high definition television. But I want to just observe a couple of trends and make a couple of my own observations.

There is a major trend away from broadcast television via the airways to cable. Cable television is roughly 30 percent of the market today and ten years ago it was hardly any part of the market.

The second observation is that the cable used to receive television signals today is intrinsically an analog medium, whereas fiber is intrinsically a digital medium.

High definition television demands higher quality displays than we typically find on televisions today, and they are more in alignment with the types of displays that we find on high-end work stations.

I believe that in 10 to 20 years the home computer will be the television. It will be tele-everything, basically, and it will receive its information predominantly through fiber optic technology.

Senator GORE. Let me just interject and say that is one of the reasons I have legislation to allow the telephone companies to get into the cable television business.

Dr. CLARK. I understand that is one of the sections.

Senator GORE. With protections against cross-subsidy.

Dr. CLARK. Excuse me, Senator.

Senator GORE. Excuse me for interrupting. Go ahead.

Dr. CLARK. I understand that is one of the sections of this bill, or implicitly——

Senator GORE. No, no, it is a different piece of legislation. I want this one to pass.

Dr. CLARK. There is some implicit deregulation in Section 202, paragraph A.

In any case, high definition television will be the technology that could yield a number of benefits. One of them is that a gigabit of network bandwidth would be required by the consumer, and the consumer market yields lots of money. I believe that by linking these in some fashion and enabling—the other aspect of what you just mentioned, namely—the regional Bell operating companies to pipe these signals into the home, it will in a sense leverage the entire technology so that we have that kind of bandwidth available at very low cost for use with computers as well.

So I consider them very tightly related.

Senator GORE. I cannot resist interrupting one more time, and I apologize. I agree with you totally, but my earlier comments were aimed at the political reality we face in Congress today, which is that the country is not yet ready to seriously consider allowing the Bell operating companies, to get into the cable television business.

I think there are ways to design protections against cross-subsidy and anti-competitive behavior that would make it eminently in the public interest. And it would be the most efficient way to wire not just the backbone network, but really the whole country, with fiber optic cable, and that is what we need to do.

But for now we have to recognize that the chances of that passing are not very good. So we need to concentrate on the backbone network, and over time I think it will become clear that the benefits far outweigh the liabilities of wiring the entire country in that fashion.

Dr. CLARK. Just one final observation in that regard. The Japanese have taught us a lesson, that the consumer market is where the leverage comes, and that is how they have achieved a dramatic improvement in their electronics technology.

I do believe that, to the extent that we as a country focus our energies or combine the objectives of this data communications network activity with the (HDTV) consumer marketplace, I think we will find greater leverage.

I have a few comments on software as well. There are many people here who have lots of experience with software, but in my reading of the bill there was too large a focus on artificial intelligence software, at least from my perspective.

I believe, as defined in the bill, some AI research is appropriate. However, I also believe that, although today you have been referring to visualization as a key element, it was not called out in the bill, as a specific line item. I think that would be an important thing to do.

Visualization software is as important as the numeric applications software that we are discussing because, as has been pointed out throughout these hearings, it is the way in which we comprehend the data that is being computed. Visualization is the key element of that.

Another element I think is networking software. It is perhaps implicit in some of the other titles, but it is not specifically called out in the title, the software portion of the bill.

In summary, again I think that this national network is probably the most important piece of legislation that has emerged in the last 10 to 15 years in terms of making our country more competitive, for all of the reasons that have been highlighted here.

I would like to reiterate, I think tremendous leverage could be achieved through combining these networking efforts with consumer technology vis-a-vis high definition television.

I also think that it would be a good idea to try to focus the bill on helping to develop better display technology. Today, all of the companies that you see selling computer equipment buy all of their CRT's from Japan. We could perhaps develop flat-panel displays or some alternative technology, and I would like to see some focus on alternative display technology called out specifically as an item in the bill.

Senator **Gore**. That is yet a third piece of legislation. We have had some hearings on that, too.

But go ahead.

Dr. **Clark**. And then finally, just to reiterate, I believe that the computing future of this country, the networking future, and the consumer electronics future are all inextricably related, and fiber optic technology is the technology that will allow those to be tied together.

[The statement follows:]

Testimony before the Senate Subcommittee on Science, Technology and Space

regarding the National High Performance Computer Technology Plan (S. 1067)

by
Dr. James H. Clark
Chairman
Silicon Graphics Computer Systems

July 26, 1989

Background.

I am pleased and honored to be asked to testify regarding the merits of the National High Performance Computer Technology Plan. I would very much appreciate, in addition, the opportunity to testify to the appropriate subcommittee regarding industrial competitiveness, as I believe even more urgent legislation is needed to improve the economic foundations of American Industry relative to the Japanese.

Before making specific comments on this Plan, I wish to emphasize that I am a strong proponent of government funding for research projects on advanced computer and electronics technology. From 1979 to 1982, as a professor at Stanford University, I was one of several Principal Investigators of a DARPA contract. DARPA funded this research to promote development of Very Large Scale Integrated Circuit (VLSI) design tools at Stanford, MIT, Caltech, Carnegie Melon and UC Berkeley, so that the United States might retain its leadership in semiconductor design. Three projects from my research contract specifically at Stanford have developed into three major companies in the computer industry:

1) SUN Microsystems was started to commercialize the SUN Terminal developed for the Stanford University Network (SUN).

2) Silicon Graphics Computer Systems was started by me to develop the IRIS 3D Visualization Computer, as an outgrowth of some advanced VLSI circuits developed as part of our research.

3) MIPS Computer Systems was started to develop a commercial version of a class project that yielded a Reduced Instruction Set Computer (RISC) VLSI chip.

The research goals of developing VLSI design tools provided an impetus to focus smart minds on a few key problems. The technology that resulted, however, has not only yielded better design tools, but it has also created three highly competitive *computer* companies whose combined revenues total almost $2.5 billion/year, just eight years later. Moreover, these companies are growing at roughly 60% per year. About 40% of their revenues are outside the U.S., helping to offset our balance of trade problems. No amount of legislation could have guaranteed the outcome, and no one predicted the formation of three excellent companies, but the funding provided fertilizer for the growth of vital competitive advantages for our country. I see the Plan before you as an even more important funding effort than this VLSI design effort of the early eighties.

I have been asked to comment on all aspects of the National High Performance Computer Technology Plan (Plan), with special emphasis on Title IV - Software.

Title II - National Research and Education Network.

This is the very backbone of the entire Plan. In the last ten years, computing technology in the most widespread computer systems in use in the world has advanced in speed by a factor of almost 60, while the cost has been reduced by a factor of almost 10 for comparable computing resources. This is a combined factor of 600. Yet during the same period, the most widespread digital communication mechanism between computers, Ethernet, has hardly advanced at all -- by about a factor of 3. Common computing equipment routinely performs at internal data rates that exceed intercomputer communication rates by a factor of more than 200. Fiber-optic technology is the key to bringing these two rates more into alignment, but the technology is still too expensive to be commonly used.

Fiber-Optic Data Highway -- Is 3 gigabits per second sufficient bandwidth?

Senator Gore has referred to the fiber-optic network as an "Interstate Data Highway System," and this is an apt description. The specific proposal is that a 3 gigabit per second fiber-optic digital network be established for use by government, industry and education. My only concern about this proposal is that it is not ambitious enough.

Consider the following data:

1) A single static HDTV (High Definition Television) image contains roughly 1/30 of a gigabit of information, so that the broadcast of only one digital HDTV channel requires a bandwidth of about 1 gigabit per second (30 images per second).

2) A compact digital audio disk holds about 1 gigabit of information.

3) AT&T Bell Labs has research equipment that achieves 200-300 gigabits per second.

4) Modern personal computers, such as a Macintosh SE or an IBM PC, routinely manipulate data at rates of 1/30 gigabit per second.

5) Ordinary $10,000 workstations equipped with RISC microprocessors have internal data rates of roughly 1/2 gigabit per second.

6) The most powerful Visualization Supercomputers available *today* (priced in the $100,000 range) routinely manipulate data in the 2 to 5 gigabit per second range.

Twenty to forty personal computers, four workstations or two Visualization Supercomputers made with *today's* technology could easily saturate a network with only 3 gigabits/s bandwidth, and this network is to span the country serving thousands of individual computing nodes. I believe that a 3 gigabit per second "data highway" could very well be a "country road" by 1996.

High Definition Television could use the same technology.

The U.S. has lost the consumer electronics industry to Japan, but it maintains a commanding lead in both the computer and networking industries. The home computer of ten to twenty years from now, however, will use the same electronics as the HD television, telephone and tele-everything else. Moreover, virtually all of the information will be delivered to it via fiber-optic digital data links. For this reason, it is very important that the U.S. leverage it's knowledge of high-speed digital communication into the home consumer marketplace as quickly as possible. The Japanese win economic wars because they understand market volumes, and the consumer marketplace yields the most volume and therefore much lower costs and higher leverage.

There are several important observations:

1) There is a major trend away from broadcast television to cable and satellite television -- ten years ago, cable was insignificant, yet today it has more than 30% market share.

2) Cable is intrinsically an analog medium, whereas fiber-optics is an intrinsically digital medium -- fiber-optics is the audio CD equivalent for communications, whereas cable is the vinyl disc equivalent.

3) HDTV demands higher quality displays than the glass CRT commonly used in today's televisions and computer workstations, which incidentally are exclusively supplied by Japanese companies. The technology used here will also be used for computer screens in the home and in the work place. Considerable research dollars should be spent in developing alternative display technology, such as flat-panel displays.

4) The technology required for HDTV is already at least 80% in common with the technology of advanced computer workstations.

Debating over how to allocate analog broadcast transmission frequencies for HDTV is not nearly as meaningful as enabling high bandwidth fiber transmission into the home, where consumer volumes will automatically precipitate price reductions that leverage computing technology as well. My view is that the airways will not be significantly utilized to broadcast HDTV, because the extra frequency spectra required will conflict with other broadcast needs, such as portable telephones, existing TV broadcasts, and so forth.

Supercomputing and Distributed Supercomputing.

Supercomputers need to be made available to a wider audience, and establishing the National Research and Education Network addresses this issue by providing long-distance connectivity between the NSF Supercomputer Centers and potential users. Another type of accessibility is equally likely to be enabled by inexpensive, gigabit networking technology, however. This is what might be called distributed supercomputing.

Several companies, including my own, have what we call Micro Supercomputers on the market that perform at numeric computation rates between 1/4 and 1/2 that of the most expensive supercomputers on some problems, at prices in the $200,000 range. These machines are accessible because of their relatively low price. They achieve high computing rates by connecting up to eight RISC microprocessors together to collaborate on a large problem. The number of processors collaborating on the problem is limited by the size of the connectivity bus. If a low-cost, high bandwidth fiber-optic link were available, many more processors, separated by greater distances, could collaborate on a problem, thereby enabling computing rates several factors of 10 higher than those of the highest performance centralized supercomputers of today, at a tiny fraction of their cost.

Tomorrow's supercomputers will no doubt be much more powerful than today's, and no one can accurately predict whether "distributed supercomputing" will be the prevalent technology. Research dollars focused on ultra high-speed digital fiber-optic communication, however, will help both technologies develop and is biased toward neither.

Title IV - Software.

This portion of the Plan addresses the development of portable, accessible software of two predominant types: 1) numerically intensive software commonly used in scientific and engineering applications on todays supercomputers and 2) artificial intelligence software.

The relative lack of availability of supercomputers to a large body of users has resulted in a relative lack of high-quality scientific software. Part of this will be solved by making these resources more accessible over the National Network. But support dollars are still necessary to encourage such software to be developed, and to support the graduate students and other researchers as they develop this software. My instincts are that some of the dollars will have to be spent on computers and computer time. I assume that this is implicit in the Plan, but it might be worthwhile to specifically call for money to buy equipment and computer time.

Artificial Intelligence has been an elusive dream for at least three decades. As defined in Section 401 (c), I believe it deserves significant funding, but I question the emphasis placed on funds spent here. If the emphasis is necessary in order that the administration of the Bill lead to equal funds being spent in this area, I agree with it. However, I believe high-quality scientific software is equally important to the objectives of the Plan.

Two other types of software are not specifically mentioned in the plan but are equally important. The first is software for *visualization* of scientific data. Visualization is the only way to efficiently and quickly comprehend the masses of data generated in numerically intensive applications, and software for visualization is as important as the numerical programs themselves. Funds should be specifically called for to develop this type of software and purchase the visualization equipment.

The other category of software not called for in this Section is communication software to access the network. I would expect some of the funds in this portion of the Bill to be available for this, thereby leaving most of the funds in Title II for the establishment of the network hardware.

Title V - Computer Systems.

Of particular importance is Section 502 related to promoting development of new display and networking technology. I have explicitly mentioned the importance in my comments on Title II, but if the money to develop these technologies comes from this portion of the Bill, the funds left there can be used to establish the network.

I agree with the intent of Section 503, that U.S. manufacturers should be allowed to export their computer systems. We permit the "export" of licenses to Japanese companies by U.S. manufacturers of computers and semiconductors, yet we restrict the export of some of the computers themselves. The Japanese companies certainly will not restrict the sale of the technology resulting from the licenses, and in the process they generate an even bigger economic machine.

Summary.

I believe that a national research and development priority on "Ultra High Bandwidth, Low-cost, Digital Fiber-optic Networking" is extremely important to the computing and consumer electronics future of this country. The evolution of High Definition TV is intimately related to this, since such a network would be the preferred delivery means for HDTV signals. Moreover, HDTV and home computers of the future will be the same electronics. In order to leverage the development of the appropriate technologies, therefore, consideration must be given to the funding implicitly available from purchases by the consumer. This is the lesson to learn from the Japanese.

To reach the consumer, ultra high bandwidth fiber optics must be provided to the home as rapidly as possible. In this regard, the Regional Bell Operating Companies are the fastest mechanism, because they already own the infrastructure. Allowing them to provide the pathways for HDTV will potentially require regulation, since they have a natural monopoly, but I truly believe that the more quickly we establish this consumer priority, the more quickly the technology will be driven down in cost, thereby leveraging the advancement of our domestic computer and networking industries.

The 3 gigabits per second objective of Title II is certainly achievable, but the ability to reach higher bandwidths in the future might be achieved by ensuring that more fiber is put down for future use. Also, by installing or allowing for later installation of fiber and repeaters capable of higher bandwidth, future expansion is ensured. The Synchronous Optical Network standards (SONET), being defined with HDTV consumer markets in mind provide for long distance transmission of data at rates up to 13.2 gigabits/s. Perhaps this would be a more appropriate target than 3 gigabits/s.

Through lower cost, "distributed supercomputing" is as likely as centralized supercomputing to yield high accessibility to a large number of users with supercomputing needs. A low cost, ultra high bandwidth network will do as much to enable this type of development as it will to enable access to centralized supercomputing resources.

I believe that the Plan should specifically emphasize in a separate Section the development of alternatives to the glass CRT for display technology. Doing so could enable the U.S. to recapture a key element of future consumer electronics from the Japanese, as well as provide the essential means for information presentation by computer.

The futures of our computing, networking, and consumer electronics industries are inextricably linked. High-speed fiber-optic communication is the key technology that applies to all of them.

Senator GORE. In the legislation that we are talking about a lot here today, S. 1067, I will just note that Title V, where we talk about computer systems and procurement, we say: "Particular emphasis shall be given to promoting development of advanced display technology." And then it goes on to mention "very high-speed communications links."

We want the networking software not only to be implicit in the bill as a matter for emphasis, but also explicit. We will take your comments to heart and study the relative attention paid to visualization and artificial intelligence.

We are going to have to move right along now.

Let me now turn to our third witness on this panel, Dr. William Poduska, the CEO, Chairman and CEO of Stellar Computer, Inc. Welcome. Please proceed. A fellow Tennessean; we are mighty glad you are here.

STATEMENT OF DR. J. WILLIAM PODUSKA, SR., CHAIRMAN AND CHIEF EXECUTIVE OFFICER, STELLAR COMPUTER, INC.; ACCOMPANIED BY NORM SIEGEL, SYSTEMS ENGINEER

Dr. PODUSKA. Yes, sir, I was born and raised in Memphis, Tennessee. And just a brief personal note. My folks knew your dad and in fact it was a colleague, Cliff Davis, Representative Cliff Davis, who appointed me to the United States Military Academy. So it is a double pleasure for me to be here today.

Senator GORE. Great.

Dr. PODUSKA. Let me just briefly focus on a couple points that are contained in my written testimony, a word or two about visualization of scientific information, a comment about what small business, especially American small business and the venture community, can do in the software development propagation effort, and then a couple of comments about the bill itself. And we have a demo here for you, sir, if you care to take a look.

First, this business of scientific visualization is in my mind the most important advancement in software for science that we have seen in really quite some time. It has been pointed out several times that the amount of data is truly massive by almost any measure.

But perhaps an example of the way that visualization works would serve some purpose here. Consider for a moment air flowing over a wing and the problem of calculating the air flow and the parameters around it. The number of grid points associated with that, at least a recent calculation I saw was a million data points—a million grid points, rather.

At ten data per point, that is about ten million double precision numbers over a complex range of flow velocities and angles of attack and such things—surely insurmountable amounts, quantities of data for anybody to look at.

If you were going to look at it, what would you see? We cannot see pressure, we cannot see temperature or entropy or things like that.

In fact, to look at it at all what we have to do is go through a two-step process. We have to construct something to look at, then we have to look at it. In the example of the flow, maybe what we do is construct surfaces of isopressure, equally spaced say three millibars apart over the wing.

We might even color those surfaces with temperature or entropy, maybe red as hot and blue as cold or something like that. We might even try some experiments floating massless particles over to watch how the flow goes.

But whatever it is, we have constructed a model and now it is time to look at it. And we do that the same way that you and I would look at any physical object. We pick it up, we rotate it, we examine it under different lights. We turn our head and cock it around to look at it.

What we are trying to do is communicate with this computer, which is surely the best one around today for anything that is going to be decisive—deciding what we are going to build, how we are going to build it, what materials we are going to use, how the air masses work, and such. We have got to see with the eye and with the mind's eye what is going on.

So it is for things like that that you get the scientists to come up with the "Aha." You know, they look at something and they say, boy, now I know how it works or why it works, or at least I have gained some insight into it. I think it is no accident about our language that we use a word like "insight" to talk about viewing into a problem almost visually.

I think it is important to note, too, that the tremendous progress in visualization has come about recently from some tremendous technological advancements, and many of them on the hardware side. The fact is we could not economically afford to put large amounts of computation and graphics systems in front of an individual scientist or engineer just a few years ago.

But with the rapid advance of technology that we all know about, many companies are able to provide equipment that allows scientists and engineers to both compute the image they want to see from the raw data, from the data such as Karl was talking about, construct an image and then take a look at it.

That is the thing that has happened over the last two or three years which has made visualization economically feasible for scientists and engineers, at least in my judgment.

In fact, there are several companies around which have capitalized on these recent advancements and formed companies around them. Jim Clark several years back at Silicon Graphics, and a company called Ardent, and another one called Stellar are companies that provide equipment of this sort.

I would point out to you that they are all three American companies, uniquely American companies, venture-funded companies, put together by people who want to approach a new idea and go and tackle it and bring something to the marketplace.

Let me connect that idea then more with the role of businesses and venture capital and this set of areas of software and visualization and network. Software is reaching a time when it is not just a matter of programming things, that is constructing a new set of programs to do something, but one of combining programs and results that other people have put together.

The visualization effort is one where the basic calculations might be done with numbers of programs all around that result in data that can then be rearranged, restructured, re-examined, and then looked at with different sets of software.

This building block approach is one that is becoming more and more visible in the software industry. It is a divide and conquer strategy that has worked well many times before. It is something that small companies are ideally suited for. It is a set of ideas that have been promoted by many small companies and will continue to be.

It is in fact one of these areas that is nicely nondiscriminatory. Ideas come from anywhere. They are unisex, color-blind.

It is possible for anybody with the right set of ideas for constructing a software component that fits into this thing to climb out on the end of that limb, get somebody to fund them somehow, and build a company that supports the software and puts it into the environment that all people can use.

I comment here also about technology growth in general. We all know how fast technology has grown. The examples that we hear are in fact just mind-numbing. I mean, a 20,000 mile per hour automobile, one that weighs three grams or something.

But we have a view sometimes that technology just continues to grow in this one long, smooth curve. It does not. It comes in big step functions. And in fact, in the computer business and in the electronics business in general, I think the remarkable thing is how steep are the steps and how flat are the plateaus.

We do not talk about going from a one megabit DRAM to a 1.2 megabit DRAM, a 20 percent improvement. We are going to go from one to four, and from four to 16. These big step functions apply throughout the industry.

Then the time to capitalize on an idea is when it is on one of these big rises. What I would say to you today is that the area of visualization is one that is right at a step function in the technology. We are just now at a time when the equipment is reasonably economically affordable for engineers and scientists to take a good look at what the data is.

Now is the time to jump on that curve, in my opinion. And I think that the efforts of S. 1067 are ideally suited to propagating that. This is a uniquely American phenomenon of the venture company that is able to take over an idea and do something proper with it. It is one that we should support and encourage.

We should encourage our youngsters to take risks, to be gamblers, to be riverboat gamblers if you please, and be able to go out there and do something.

A word on the bill itself. I am certainly not an expert in legislative matters and I would not pretend to comment about what belongs where or such. But I can make three observations about this bill.

First of all, the national research and educational network at three gigabits in 1996 is just plain a great idea. It is supported by our experience in networking. We know that people will use it.

We know that they are limited by bandwidth right now and, as has been commented several times, three gigabits is a good goal for us to reach now, but we should be looking for ten times that by 2000 or so.

The clearinghouse for software is another idea which is just great, but it depends on the network. It does no good to just have a trickle flow of information back and forth from scientist to scientist.

And the support of visualization that is implied by the bill I think will be connected to all three of those and wisely, I think, in the words of Bill Wulf, in this notion of a collaboratory, because we can do more than just send messages back and forth. We can work colleague to colleague with each other.

Those are my comments on the structure. We have, if you please at this time, or later if you prefer, a sort of a show and tell of the visualization effort.

[The statement follows:]

TESTIMONY OF JOHN WILLIAM PODUSKA

U.S. SENATE HEARING ON BILL S1067

JULY 26, 1989

Honorable Senator Albert Gore
Honorable Senator Ernest Hollings
Members of the Senate Subcomittee on Science, Technology, and Space

Gentlemen:

It is indeed a pleasure for me to have an opportunity to testify before this Senate Subcommittee. The subjects of visualization, artificial intelligence, and advanced computer software are very dear to my heart and central to my career. They are areas in which the United States has a continuing sustainable edge in world competition.

My comments and testimony are organized as follows:

1. General Comments Regarding Software and Programming.
2. On Visualization of Scientific Information.
3. On Artificial Intelligence.
4. On Advanced Computer Software.
5. Comments Specific to S1067
6. The Role of New Business and Venture Capital.
7. Conclusions.

1. General Comments Regarding Software and Programming.

The computer industry is very young and has not developed a truly independent vocabulary yet. Thus, one must be careful to avoid applying the dictionary meaning of words to this industry. And so it is that software is not "soft", and hardware is not "hard", and compilers don't "compile", etc.

While these distinctions may seem obvious, it is more important, I think to separate the notion of software from programming. We tend to think of hardware as the things that electrical engineers build, software as the things that programmers create to run on the hardware, and programming as the knack of creating and executing interesting sofware on computer hardware. In fact, until quite recently, the notion of using a computer well was synonymous with knowing how to program well.

But, the recent progress in computer usage has come from separating and removing programming from software usage. Some examples seem appropriate here:

a: Spreadsheets - makes financial analysis possible without COBOL or RPG or DBMS.
b. Visualization - makes insight into scientific and

engineering problems possible without graphics programming.

c. Artificial Intelligence - makes deductive systems possible without programming in LISP and often with unexpected results.

I believe that the software portions of S1067 are properly aimed directly at accelerating this process, especially the high performance National Computer Network, the software clearinghouse, and support for visualization in the scientific community.

Modern software is much more than just a given program. Software elements can be fitted together like bricks in a wall if the interfaces are correct, they need not be like poured concrete. But for this to work, there must be accepted standards for languages, operating systems, graphics interfaces, and file structures. The recent successes of C, UNIX, X-WINDOWS, PHIGS, and NFS help a great deal. I believe that object oriented programming systems (OOPS) will have major benefits here and should be supported. But given the tools, we need individuals to build the blocks and share them: hence the need for the National Computer Network, and the Software Clearinghouse.

One last general comment, software development has traditionally been a risky individual effort. It seems to be much more closely matched to the American frontiersman culture than to the more traditional European or Japanese culture. Part of the reason for American success is the areas may lie in this observation.

On Visualization of Scientific Information.

There are many very difficult problems in science and engineering which generate massive amounts of data when solved on computer systems. It is not at all unusual to find that a single solution generates in excess of 10 million items of information. This amount of information cannot be understood in numerical form by a human being. But, if we could make a picture which represents this data, a human being could understand the results of his computer.

Consider the example of airflow of a wing, analysis of a modern wing might typically take 1 million grid points with 10 data items per grid point. Suppose we start with an image of this wing and add to it artificial images that the scientist can visualize. Suppose we add surfaces of constant pressure over the wing. Perhaps we can add a color to the surface which represents temperature and to this we might add the image of small particles flowing over the wing. Now we have a composite image that the scientist can see with his eyes and with his mind that will give him insight into the structure of airflow over that wing.

Visualization of scientific information is therefore a two step process. First, the scientist contructs an image from the data and second, he views that image from different angles, different lighting, different levels of transparencies, etc. The most important point about visualization is that it connects the results of the computer in the hardware with the computer in the brain. It is the human computer that will ultimately make the design decision and gain insight from the computation that is performed. Visualization is a brand new tool for bridging the gap.

Visualization hs only become feasible in the last two years with the advancement of technology in the form of graphics supercomputers. Visualization requires massive amountsof computation as well as graphics computer power dedicated to a single scientist to achieve any worthwhile results.

Visualization is widely regarded as the most important advancement in scientific computing in the last decade. I concur.

3. Artificial Intelligence.

The term Artificial Intelligence is generally intended to mean software systems which mimic or aid human intellectual processes. The promise of artificial intelligence is very real. One can imagine expert systems and knowledge based systems which aid human beings in analyzing collections of imprecise data. Examples ranging from medical analysis to remote robotics seem feasible in some reasonable near timeframe.

But, the progress in this area has been painfully slow. Many noted scientists believe that substantial progress in artificial intelligence is not nearly a matter of increased computing power, but will require more fundamental breakthroughs in the science of computing. It may very well be that real progress in artificial intelligence will require decades of effort.

Therefore, I believe that artificial intelligence can benefit from basic and fundamental research such as is supported by the National Science Foundation. It can also benefit from the collegial environment set up by the facilities of S1067. However, I believe that artificial intelligence would be an inappropriate domain for massive funding of specific projects.

4. Advanced Computer Software.

Recently, there seems to have been a major accelleration in the development of advanced computer software in the scientific and engineering world This is in part due to the wise and timely support of the National High Performance Technology Plan. But, it is also due in large measure to the recent scientific breakthrough in areas as diverse as molecular modeling and mechanical computer aided design. It is almost as if we were suffering from a log jam in computer software and the key log has been removed.

The process has been aided by standards in the software industry as well as a recognition of the virtues of a building block approach. The result has been that individual contributors have the means to communicate both the progess and the results of their colleagues.

Whatever the reasons for this surge in advanced computer software, it seems clear that the time to exploit it is now. The U.S. is in a unique position to exploit advances in advanced computer software for science and engineering. We have an installed base of high performance computers and visualization equipment. We have a community of scientist who are willing to take professional risks. We have a community of entrepreneurs and a venture capital structure who are willing to support risky new efforts in software development. Government can foster this development by making software packages readily available in the public domain by establishing networks which allow scientists in research and industry to quickly communicate software results and by establishing a set of clearinghouses for fostering the development and distribution of new software building blocks. In my opinion, S1067 part Title 1 and Title 4 are effectively directed at this effort.

5. Comments Specific to S1067.

Senate Bill 1067 seems to me to be a comprehensive and well written document that properly and effectively addresses many of the needs of the scientific computer community. However, it is beyond my level of experience or expertise to comment on matters of structure, jurisdiction, or timing of its components. Nevertheless, my observations, as a neophyte observer, may be of interest to the committee.

First, the very notion of having a National High Performance Technology Program implemented through the Federal Coordinating Council for Science, Engineering and Technology properly places a focussed spotlight on U.S. efforts to develop high performance computer technology. The FCCSET council is well received in the scientific and industrial community even at this early stage in its development.

Second, the establishment of the National Research and Education Network in 1996 is a bold and appropriate step. Such a network will be sufficiently robust to allow large groups of scientists and engineers in all components of our society to freely exchange software images and results on a colleague to colleague basis. Network performance at 3 Gbps will be adequate in 1996 but only barely so. We should anticipate a network of ten times that capacity by the turn of the century.

Third, the establishment of the National Information Infrastructure as well as various software clearinghouses will go a long way towards organizing the exchange of information on a colleague to colleague basis and will effectively use the National Research and Education Network. In fact, the clearinghouse concept would be quite weak without a strong network to support.

Fourth, the several National Science Computer Centers in the U.S. have served the scientific and engineering communities quite well. These centers support the U.S. computer industry very well and effectively encourage small U.S. companies to develop new hardware and software products. These centers should be encouraged to continue to buy equipment and software from smaller and newer companies rather than accept a gift from the larger and more established companies.

Finally, support for basic research and education is a crucial and fundamental element for long term success in the computer business. But, basic research does not mean unguided research. In my experience, most research benefits greatly from being specifically goal oriented. There are some areas where goals and project structure may not be appropriate. Artificial intelligence may be one such area. I believe that efforts in visualization, object oriented program systems, graphics, and parallel processing can very well be goal oriented and project structured. As one example, I am particularly impressed by the center for parallel processing established by the National Science Foundation.

6. Role of New Business and Venture Capital.

The field of computer science has, without doubt, been the fastest growing area of knowledge that the world has ever seen. Most new areas of effort have come from a technology push rather than from a marketplace pull. Indeed, many products that we view as essential today were not so much as conceived ten years ago.

A good example is todays use of spreadsheets. Such technological development is a fertile field for the development of new businesses. New businesses can move more quickly and take a bigger risk than established businesses. They are well suited to the culture and personality profiles of entrepreneurs and venture capitalists. It seems natural, therefore, that so much of the new technology in computer science should be commercialized by new companies which are venture capital funded in the U.S.

I believe that this is an area where the U.S. can take justifiable pride in the way our system of small company capitalism has worked so well. It is important to remember that this culture is uniquely American and that we have an opportunity to exploit the entrepreneurial spirit of U.S. scientists and engineers. On a larger scale, we must encourage the gambler's instinct in our venture capital community and among our entrepreneurs. We must continue to allow people to fail without being condemned for it.

7. Conclusions

It is very much an understatement to say that computer science has been the fastest growing area of science that the world has ever seen. Perhaps I can put that statement in perspective for this committee by pointing out that a Stellar Computer which sells for $100K has more computing capacity and more storage than the sum of all the computers on this planet when I was in graduate school in 1959. Credit for the growth of this industry can be shared by many people, including the government, the academic world and the venture capital world.

It is equally true that continued success will likely come from continued support from the government, the academic world and the venture capital world. S1067 speaks well for the commitment from the government to continue its support. Support for the academic community through this bill and others will continue to encourage academic support for computer science. And, continued support for the uniquely American world of entrepreneurs will ensure the continued support of the venture capital world.

To quote Al Jolson from his movie, the Jazz Singer, "you ain't seen nothing yet"!

Senator GORE. How long is it?

Dr. PODUSKA. Well, we can make it as short as you like, sir. A few minutes if you prefer.

Senator GORE. Let us do a couple minutes of it now.

Dr. PODUSKA. All right. We have Norm Siegel here to do that. The machine was loaned to us, by the way, by Dr. Richard Feldman of the National Institutes of Health at Bethesda, who is doing a great deal of visualization work on molecular modeling.

Speaking of gambling, it is something to gamble to just pick it up and move it here and see if it works.

Senator GORE. Right.

Dr. PODUSKA. But I think it will.

Senator GORE. We better do it now before the thunderstorm knocks the power out.

Oh, is this it here?

Mr. SIEGEL. Yes, this is it. One thing to note is that there is a very wide array of applications——

Senator GORE. Can you introduce yourself.

Mr. SIEGEL. Yes. My name is Norm Siegel. I work for Stellar Computer as a systems engineer.

Senator GORE. Okay.

Mr. SIEGEL. And what I am going to do is demonstrate some of the capabilities of one of the Stellar products that are available today. What we are showing here is an array of possible application areas that are capable of being visualized on the system.

There is sort of a blueprint here showing various functional application areas. And what I will do quickly is go through some visualization that represents some scientific visualization application in the area of meteorology. What we are going to see here is a thunderstorm that developed over a 36-hour period over the United States.

Senator GORE. Pretty appropriate.

Mr. SIEGEL. Okay. Now, talking about volumes of data, this data set is a small fraction of the amounts of data that are available to meteorologists, and this data set that we have here today takes up a large fraction of the storage capacity of the system.

It is just an indicator of the need for high capacity systems and the ability to process that data quickly.

Senator GORE. Are you waiting for it to find the——

Mr. SIEGEL. Yes. The data sets are very rich and complex.

I will first bring up an image of the United States and move it around a bit here.

It will show some of the winds that are traversing the United States at the time of the simulation. And this data was actually simulated from data that was collected and then run through a scientific model of the weather.

Okay, and there is information here right now about horizontal winds, but I can at the same time look at other information, such as cloud water, during the course of the storm over this 36-hour period. We can also look at vertical winds as they occur.

The other interesting thing here is that we can request the system to select out different aspects of the data, to selectively view various portions of it and view how things are interrelated.

In fact, one of the developers of the scientific model was able to sit at this system and view the model as it was progressing and gain insight into improvements for the model as the result of that in a very short amount of time.

Dr. **Poduska**. Of course, that is the whole purpose of modeling.

Mr. **Siegel**. Right.

Dr. **Poduska**. In the interest of time, that might be all you want to see.

Senator **Gore**. Yes, we probably better. Thank you very much. If we have time at the end of the testimony, I may want to look at some more of that.

Thank you very much.

I appreciate the patience of our other witnesses, too. We are going to go to our final two witnesses now. Dr. David Nagel is Manager of User Technologies at Apple Computer. You all just bought a Cray, right?

Dr. **Nagel**. Well, a couple of years ago.

Senator **Gore**. And you are using it to help design better products, and we want to hear from you about that.

Please proceed.

STATEMENT OF DR. DAVID NAGEL, MANAGER OF USER TECHNOLOGIES, ADVANCED TECHNOLOGY GROUP, APPLE COMPUTER, INC.

Dr. **Nagel**. Thank you. Apple and I very much appreciate the opportunity to come here today and talk to you about advanced computing and about your bill S. 1067.

In my written remarks I paraphrased a story which I read recently in the "Wall Street Journal". The Journal did an interesting experiment about 20 years ago, which they have reported recently, in which they asked various experts to predict what technologies would be like in the year 2000. I think the interesting thing about those predictions was that most of the experts at that time fairly dramatically overestimated progress in the areas like transportation, climate control, and other things, but signally underestimated performance in the technology, information technology industry.

Perhaps that general underestimation by our society of progress in computer technology explains why in your view government policy has been so slow in responding to it. But at any rate, the pace of technology in computer and semiconductor industries is really quite unprecedented in human experience.

Various other speakers here have described this pace in various ways. My favorite statistic is that in the last 30 years there has been a six order of magnitude, or a factor of a million to one, in price-performance. And maybe even more interestingly, it is the only industry that I know of in which one can choose to take that benefit in either price or performance dimensions.

The Federal government in the past, through DARPA and the National Science Foundation and NASA and other agencies, has played a major role in all of these developments, and I think as a result advanced computers and advanced computing have really become absolutely indispensable tools in science, education, and increasingly in worldwide industry, as you pointed out in your opening remarks.

At Apple, for example, we use supercomputers both to conduct applied research in our Advanced Technology Group, where I am, and also to design and develop leading edge personal computers. Our Cray XMP/48 supercomputer is, for example, routinely used to simulate performance of very large integrated circuits, active elements numbering in the hundreds of thousands or even millions in some cases.

These cost-effective, time-saving simulations allow our engineers to make their mistakes, and design changes in software before they commit those to costly fabrication steps.

They also play other important roles in our computer development, some of them not so intuitive. The molding processes, for example, for the complex plastics that we use in our cases, and in particular in our newest Macintosh computer, the IIcx, were completely simulated and optimized on our Cray, using computational fluid dynamics codes which originally were developed in the aerospace industry.

Again, these simulations save us both time, which I think is the most important benefit, and eliminate costly redesign steps.

I actually have a very short video, which at the end of my talk if we still have time I can show you.

Senator GORE. It is just two or three minutes, is it not?

Dr. NAGEL. Yes, it is two or three minutes. So I will save that for the end if that is okay.

What I would like to do now is to turn to the major points of the written testimony, which really have to do with software parts of S. 1067. Increasingly, good software is the key to success in the usefulness of computers in the computing industry.

As I noted in my written testimony, progress in software development or software engineering has been much, much slower than that for hardware, and several people have already commented, as well as yourself, on the slowness of the increase in productivity in software development.

It is very hard to index that performance in some ways, but probably an optimistic estimate is since we began writing programs we have not seen much more than an order of magnitude improvement, or about a factor of ten, in contrast to the six orders of magnitude in hardware.

In the personal computer industry, the availability of low-cost and very high utility software, where utility is measured in terms of the user, has been really absolutely essential to our success in both this country and elsewhere. It is interesting, I think, that low-cost software for personal computers results not from significantly better methods of development than for scientific computers, although our company, Apple, and other computer companies go to great lengths and great pains to provide developers with efficient software development tools, as good as we can do at least, but low-cost software in personal computers really results because our markets are so large that development costs, large as they are, can be spread across many users.

Another success factor in the personal computer industry has been the trend toward providing powerful and very easy to use software that is customizable by the individual user. Spreadsheets, for example, which Jack Schwartz mentioned, really represent a new form of software, at least new several years ago, which enable, for example, a financial analyst to develop complex financial models without having to be a skilled programmer.

I think Jack actually pointed out one thing that is important is that the makers of these, the writers, the builders of these software programs, have gotten very good in the personal computer industry, because we have to service a large market, at matching the mental model of the user. And I think that was a very important point and one that could be of benefit in the development of advanced scientific software. It has begun to in limited degrees.

These non-procedural software languages on which our multi-billion dollar personal computer industry has been developed over the past decade are really only recently finding their way into advanced scientific computing. Programs, for example, like Macsyma or Mathematica represent those sorts of trends.

Another change that is taking place in advanced computing which will make continued development of new software methods and concepts even more central—and again, this has been mentioned by other speakers, but that is the trend toward massively parallel architectures. Commercially available computers already exist which employ thousands of individual processors and boost computing throughput to the tera-ops, or million million operations per second, level.

Billion bit per second networks will make possible vast heterogeneous computing systems with theoretical computing speeds even higher than that.

This mode of computing, by the way, has already been demonstrated outside the laboratory. It is clearly possible.

The problem is that we really do not yet have general methods of building software that can take advantage of these new hardware architectures. In more specific terms, we think that the following kinds of software progress will be needed to achieve the full benefits of massively parallel computing architectures, ultra-high-speed fiber networks, and large information stores.

New software concepts and architectures will have to be invented to coordinate action efficiently across these diverse elements and machines and to solve some very complicated problems of resource allocation and security, and both Jack Schwartz and Bill Wulf talked about some of these, particularly the security problem, which is indeed a very difficult one.

New methods of data encapsulation and exchange protocols will be required to free scientists and others from the often onerous task of reformatting and transferring their data between disparate machines. And this is something that I think is one of those chores in software engineering that is not very attractive and not very exciting, but if we really want to make productivity gains at the level of the individual scientist one that we are going to have to pay attention to.

To make very large information stores and data bases useful will require us to devise better ways to access them, and here ideally by content and not just address. Continued development of better high-level programming languages, particularly those with dynamic properties, and of better methods for software development in heterogeneous computing environments, are also important.

Finally and perhaps most difficult of all, I think we are going to have to continue to work on the problem of the human interface to these new and complex computing architectures. Along these lines, a great deal of progress has been made in the past few years in the area of scientific visualization, which a number of my colleagues have already talked about as well as yourself, and advanced methods for graphically portraying large data sets.

We will have to improve and extend these methods if the future scientist and engineer is to be enabled to spend the greatest fraction of their time doing science and engineering rather than managing the computing process.

I think one thing that is notable in this regard is that visualization has been a technology, and I think we really need to convert it into somewhat more of a science at the same time we improve the technology. It is the case that knowledge of how the human visual system works can be helpful in that regard.

And I think some of that is already being seen in attempts to find better ways of compressing information, for example in high definition TV signals and other kinds of dynamic visual images, compressing them so that when you throw away information in order to achieve the compression you do so in such a way that the visual system is insensitive to it, so the picture looks just as good.

That requires that you explicitly understand how the human visual system works and can take advantage of that in designing your algorithms.

This completes my prepared remarks. I comment more extensively in my written testimony about individual parts of S. 1067. We very strongly support it. We think that most, if not all, of the sections of it are very important. And I would be happy to answer any questions.

[The statement follows:]

STATEMENT OF DAVID C. NAGEL, PH.D., ADVANCED TECHNOLOGY GROUP, APPLE COMPUTER, INC.

Introduction

About 20 years ago, the Wall Street Journal published predictions by various pundits about what life would be like in the year 2000. For the most part, the predictions were hopelessly optimistic: domed cities rising on desert floors, hypersonic aircraft plying the airways between New York and Tokyo in two hours, and human travel to the planet, Mars. In one category, however, the reverse was true. By the year 2000, it was predicted in 1967, there will be more than 220,000 computers in the U.S. Today, more than a decade from the Millenium, there are approximately 45 million computers in the U.S. The predictions were low by more than two orders of magnitude.

The development of information technologies - computers, software, and the means to connect them - is unprecedented in man's experience. As Professor Fred Brooks from the University of North Carolina has noted, "...no other technology since civilization began has seen six orders of magnitude in performance-price gain in 30 years." Dr. Brooks goes on to point out that, in addition, in no other technology can one choose to take the gain in either improved performance or in reduced costs.

Surprisingly, the rate of performance gain is accelerating in some areas: microprocessor speeds are increasing at the rate of 70% per year. Previous generations of computers based on less integrated technology enjoyed performance improvements of less than 15% annually. Since the 1,000 bit memory chip was first developed in 1972, memory densities have increased by a factor of four every three years. Magnetic disk densities have increased by a factor of two every three years. Gigabit per second network technology based on fiber-optics have been demonstrated in the laboratory and thus should be possible to have such technology operational well before the year 2000.

In one area of information technology, progress has not been nearly as rapid: software. Although it is more difficult to index performance in the case of software, it is clear that improvements of productivity lie somewhere between a factor of 10 and 100, that is, one to two orders of magnitude. The largest improvements have been those associated with the transition from machine-level or assembly languages to the so-called high level languages introduced beginning about 30 years ago. The most famous of these is Fortran, invented by John Backus and associates at IBM, in the early 1950's. Although it has been improved somewhat since its introduction, Fortran retains most of its early procedural language characteristics. Somewhat surprisingly, Fortran is still the most widely used high level language for scientific computing (surprising since, if this practice were paralleled in hardware, many scientists would still be using an IBM System 360 as their hardware platform of choice).

High-level languages, along with time-sharing operating systems (which replaced "batch" systems), and unified computing environments (which typically include code editors, program integration tools, code libraries and file systems, debugging systems, and inter-program communication facilities), all have helped programmers to develop software faster than before their introduction. Again, the gains have been modest, particularly in comparison to the advances associated with hardware.

To attack this problem, new generations of productivity-enhancing software engineering tools are being developed and tested: new high-level languages (e.g. Ada, CommonLISP), new approaches to programming such as object-oriented or visual programming styles, computer-aided software engineering (CASE) tools, automatic programming systems, and artificial intelligence. The latter is more noteworthy for what it has failed to accomplish than for what it has, particularly in the domain of software engineering. Despite very aggressive claims that were made in the mid-60's for the general strides that would be made by that intersection of computer and cognitive science known as artificial intelligence, or AI, and significant federal funding from DARPA and other agencies, the mid-80's have come and gone with many of the claims unfulfilled. One of the most promising of the technology spin-offs from several decades of AI research, expert systems, has seen some application to software engineering: automated checklists for bug detection, optimization hinting, advice on testing strategy, style rules, and so forth. But these tools, as well as the others listed above, deal primarily with the details of programming, and not with the building of the underlying structure or conceptual basis for a program. Even with the relatively simple computer architectures of today, it is at the conceptual level that the majority of the programmer's resources are spent.

Because so much of the task of software engineering is at the conceptual level, the prognosis for dramatic increases in productivity (corresponding to hardware improvements) seems poor. On the positive side, this characteristic will make it very difficult to apply the manufacturing process technologies, which many of our international trading partners have used to advantage to dominate the world-wide semiconductor industry, to achieve equivalent domination of the software industry where the U.S. still leads. Overall, advances in computing hardware performance will to continue to outstrip advances in computer software productivity by a significant degree, although steady progress is being made in the latter. I will return to this issue in a later section.

Applications of Advanced Computing Software

Scientific Computing

Less than 10 years ago, there existed significant debate about whether computational science could in any way be considered co-equal to theoretical or empirical science. The debate centered on the essential question of whether computers were mere tools for generating and analyzing data or whether the use of computational methods could yield results and insights qualitatively different from those provided by the more traditional methods of theoretical analysis and experimental verification. As far as I can tell, that debate has largely subsided in favor of computational science. In a number of instances over the past few years, the use of advanced scientific computers has led to fundamental new understandings of phenomena at scales ranging from the structure and behavior of the fundamental particles, to the flows and dynamics of real fluids and gases, and the origins and structure of the universe. The federal government, through the National Science Foundation's Advanced Scientific Computing Program, which established a system of advanced computing centers at major institutions in the U.S., has played a major role in these developments. Even earlier, DARPA's pioneering work in networking and VLSI design and fabrication technology developments led to many of the foundations on which modern computer science and technology are based.

Industrial Computing

Supercomputing has also become an important tool for U.S. industry, particularly aerospace. For example, a fairly complete theoretical basis for computing the flows and pressure fields around airfoils and aerospace vehicle bodies has existed since early in the century. Recent advances in computationally based turbulence models have improved the power of these theories and have provided the necessary foundations to compute complete flow fields and to optimize vehicle designs as a result. As a notable example (and one which highlights the productive relationship that exists between national facilities, in this case the Numerical Aerodynamic Simulation facility at NASA's Ames Research Center, and the U.S. aerospace industry), Boeing recently redesigned the engine nacelles for its 737 aircraft. Although earlier wind tunnel testing had ruled out a proposed change as inducing too much drag, computations at the NAS supercomputer complex showed that small modifications of the basic design would produce acceptable drag performance while meeting other design objectives. The new design was verified in further wind tunnel tests and implemented by Boeing.

Apple Computer provides other examples of the benefits of advanced computing to U.S. industry. Apple operates a Cray XMP/48 supercomputer as part of its engineering complex. This high speed computing system is routinely used by Apple to simulate the performance of complex integrated circuits with elements numbering in the hundreds of thousands or more. By simulating these circuits before designs are committed to costly fabrication steps, Apple saves both time and money.

Recently, Apple began shipment of a new and highly popular version of its Macintosh II computer line, the IIcx. Most notable in the present context is the fact that the design for the complex plastic case for this computer was optimized through the use of mold flow simulations run on our Cray supercomputer. Again, simulation of the mold dynamics prior to the construction of very costly injection mold forms resulted in both a superior and more defect-free product as well as reducing the total time to market for the overall computer system by reducing the need for costly re-designs.

Finally, we use our supercomputer at Apple to conduct basic scientific and technical research. Richard Lyon and Malcolm Slaney of our Advanced Technology Group, for example, have recently developed models of the functional dynamics of the human hearing mechanism. These complex models, which would be entirely impractical for execution on anything less than a supercomputer, may eventually form the basis for advanced speech recognition systems for personal computers and improve dramatically the performance of human interface designs with which Apple has become synonomous.

The Importance of Software

Although our ability to design and produce personal computing systems has been propelled to a great degree by the increasing availability of high density, low cost memory and ever more powerful microprocessors that represent the raw computational power of the largest mainframe computers of the past decade, it has been the widespread availability of low cost software that has fueled the personal computer revolution.

Personal computers have become so important in our lives because they have become so valuable to us as as the means to increase personal productivity and improve our ability to communicate complex ideas. The systems are both powerful and, in many cases, relatively easy to use. Whereas prior to the explosive growth of the personal computer industry, individuals' usage of computing systems required a signficant personal investment in learning to program and to master the arcana of the "computing center," personal computers have placed high performance modelling, word processing, and database manipulation capabilities in the hands of millions. Software components which implement new forms of human-computer interfaces that are more intuitive, require less learning, and and that can be customized to suit individual needs, have played critical roles in the successful introduction of computing to the mass market. Because they are both powerful and easy to use, the utility thus provided by personal computers has, to an increasing degree, made them a virtual necessity for many professionals engaged in competitive work situations. Thus, very powerful re-usable software codes for personal computers which, because of the very large installed base of computer systems (numbering in the millions of units in the U.S. alone), can be sold for prices ranging from tens to a few hundred dollars.

It is not the case that such software can be developed more quickly than systems for use on larger computing systems. The software engineering problem has not been solved entirely by the personal computing industry although significant resources within the industry are routinely committed to improved systems to improve and speed the development process. Rather, it is the combination of a large market size and the close relationship which Apple and other personal computer manufacturers maintain with 3rd party developers which support a low market price. Since the development costs for any particular program (which may easily reach many millions of dollars) can be spread across a total market size numbering in the hundreds of thousands or millions, the portion of this cost which must be borne by individual users of the software is reduced to a very low level. The personal computer industry provides a vivid example of the benefits of powerful, easy to use, and high utility software systems.

In the high performance computing world, the importance of software will become even more apparent as new computing architectures are invented to deal with the physical limitations of single processor systems. Earlier in this decade, it became apparent that there were ultimate limits in the power of any single computing engine, limits provided by the speed with which individual computing elements can change state, by the speed of with which signals traverse the internal data paths of processors, and by mechanical and thermodynamic problems caused by the large amounts of heat produced by very fast computing circuits. Since then there has been increasing interest in computer architectures which employ numerous processors performing in parallel with one another. Although the number of processors may range from a couple to thousands, it is clear that we do not yet know how to build software systems that can take full advantage of these new parallel architectures. For example, experience at NASA has shown that user codes sustain only about 20% of the full power of a Cray-2 supercomputer, a moderate grain parallel computing system with a small number of very fast processors. Experience on more massively parallel systems such as the Connection Machine produced by the Thinking Machines Corporation has produced prototype codes which sustain only about 10% of the peak power theoretically available. While these percentages can be increased to nearly 100% for particular computing problems with the proper attention to code formulation and optimization, such human interventions significantly reduce the overall utility and productivity of these powerful computing resources.

Thus, although a significant success factor in personal computing, computer software shapes up as the major bottleneck impeding the full realization of the power of massively parallel computing systems of the future.

New Paradigms in Computing

The dominant computing paradigm of the past decade (time sharing operating systems, large central processors accessed by relatively low power alphanumeric computing "terminals," networks with speeds in the thousands of bits per second range, and results of computations expressed as tables of numbers) is in the process of being replaced by a new one.

First, networks are both faster and more universal. Stimulated initially by the DOD's then Advanced Research Projects Agency's (ARPA) high speed computing network (Arpanet), the evolution of universal computing networks has been paced by: continuing government investments in computing networks (CSNet, NSF's ScienceNet, and Internet, to which the Arpanet has evolved); the growth of informal networks such as the Usenet (a worldwide news and mail system supported by nearly every machine the runs the Unix operating system; a host of commercial networks and mail systems such as MCI Mail; and gateways which provide various kinds and qualities of interties among the various transport systems. Fast scientific networks operating in the speed range of tens of thousands to a million or more bits per second have begun to make practical the transport of a different form of data between computing nodes: images. Viewed in one way, the "worldnet" has become a vast heterogeneous computing system, in which a single user may theoretically gain access to a vast array of computing resources (sometimes to pernicious ends). The new networking capabilities have resulted both in increased choice of computing resources for the individual scientist (or increased complexity, depending on how one views it) and in increased potential for effective scientific collaboration over long distances.

Powerful personal computers and workstations, individually capable of significant levels of computing performance, are rapidly replacing "the dumb terminal." Finally, the dramatic increases in computing speed available through individual processor performance improvements and moves toward "massivel parallel" computing architectures have encouraged individual scientists and engineers to consider and tackle problems of increasing scale and complexity. Computational results in the form of numerical tabulations are no longer useful when data entries number in the millions. Humans simply cannot perceive patterns in such datasets. New methods of representing the data, including advanced graphical or visualization techniques, increasingly characterize the form in which scientific and engineering data are presented to individual users at remote computing nodes.

In this new computing paradigm, fueled as it is by ever more capable hardware systems, software will again turn out to be the dominant factor in the overall productivity of the computing system. As evidence of the importance Apple places on development of software that will maximize the productivity of individual scientists, we recently entered into an active partnership with the NSF-sponsored National Center for Supercomputer Applications (NCSA) at the University of Illinois to support the development of scientific computing software which: (1) allows individual scientists to access remotely the full power of supercomputing resources at NCSA and other insitutions, (2) supports the interactive visualization of complex datasets generated on supercomputers, and (3) facilitates the efficient interchange of data among a diverse and heterogeneous computing network, all from the scientist's own personal computer. Increased availability of battery operated portable personal computers over the next few years will free the scientist even from the confines of his or her own office. All of the software which has been produced as a result of this partnership between Apple and the NCSA has been placed into the public domain so that it may benefit the largest number of potential users and, as well, so that it can demonstrate the benefits of new personal computers as cost-effective scientific computing resources.

Software to Support the New Scientific Computing Paradigm

Based on sustaining the current rates of development, we can assume that by the year 2000:

- The fastest computing resources will exhibit computing speeds in the range of "tera-ops" (a million-million operations per second);
- Primary memory systems will exhibit storage densities in the gigabit (billion) range;
- Information will be transferable between all machines at gigabit per second rates (enabling routine transport of video, complex graphics, and sound, as well as numbers and symbols);
- Individual ("personal") computers will exhibit performance at or above the lev

els of today's fastest supercomputers, and;
- A signficant fraction of human knowledge will be accessible electronically.

How will such computing resources be used and what kinds of software developments will be required to realize them?

First, it seem clear that collaboration will become a significant, if not dominant, form of scientific activity. Once scientists can interact with colleagues visually, auditorialy, and through use of shared computer-mediated workspaces, they will be able to work as efficiently as though they were co-located. Second, they will wish to have the ability to gather instantaneous computing resources and power commensurate with the problem being solved, which may exceed the processing limits of any individual machine. Third, they will seek increasing access to stored information for purposes of scholarship and to flexible and comprehensive mathematical modeling environments. Finally, scientists will utilize local computing resources (the "personal computers" of the next decade) to perform a significant fraction of the functions of modeling and problem formulation, data handling, visualization and understanding, and the development of materials (electronic and otherwise) for communication with their peers.

To support this scenario, significant developments will have to take place in software science and technology. Software for a complex network infrastructure will have to be developed and maintained. This will involve solution to problems of coordination of action across the diverse computing machines, to problems of resource naming and topology, and to problems of security and privacy. New methods for accessing immense stores of information and for determining the content of these stores will also be required. Robust methods of data exchange and of coordination of computing code components residing - and executing - on physically separated computing resources will require development. Finally, and perhaps most challenging of all, improved methods of coupling the individual scientist or engineer to the overall distributed computing system will have to be invented. Even with the promising developments of the past several years, far too much of contemporary scientists' time is spent on dealing with the computing system rather than with the scientific problem at hand.

The Federal Government's Role

The government can continue to play a significant role in the development of such a national capability although it is clear that industry can and will provide a significant fraction of the critical technology components. First, federal funding can provide a stimulus or "jump start" for technology development. As well, the government can serve as a "first customer" for high risk products, particularly those produced by the small innovators where many of our best new ideas emerge. The government can play an important role in development of the network infrastructure including both physical transport and network management software. The government can and should support re-usable scientific computing codes, ideally written in modern computing languages to facilitate modularization and inter-operability across diverse computing resources. The development of data exchange architectures and protocols can also be facilitated by government intervention, although industry will play and important role here as well. The government must continue to have a strong role in the education of computer scientists and those in other disciplines and to continue to ensure that the U.S. system of higher education has available the state-of-the-art in computing resources. Finally, I believe that the government has a legitimate role in monitoring and analyzing the societal effects of advanced computing and to encourage developments that will best serve the achievement of national social and economic goals and the improvement of the welfare of our citizens. In particular, the application of advanced computing resources to solution of problems in the economic, life, and behavioral sciences should strongly be encouraged.

Specific Comments on S.1067

Title I: "National High-Performance Computer Technology Plan"

A comprehensive plan for a program of this scope and magnitude is an obvious necessity. I would urge, however, that simplified and efficient means be developed for inter-agency coordination and for assessment of the accomplishments of the Act. Specifically, I would recommend that the Congressional Office of Technology Assessment be asked to conduct periodic studies of the effects of the conduct of the Plan and recommendations as to ways to better leverage the government's investments.

Title II: "National Research and Education Network"

I strongly urge that in the implementation of this network signficant attention be placed on ways in which to improve the specific ways in which people interact with such a network and on provision of systems and protocols which support the dynamic allocation of computing resources to solve problems of varying complexity. Development of the science and technology bases to support such a mode of operation must take place prior to the implementation of the network, of course.

I am also very concerned that the federal government maintain sight of the necessity to support computing in the commercial and private sectors. The current regulatory climate makes it overly difficult for private enterprise and the utilities to install very high performance networks to support business, consumer, and local educational institutions. Although it is clear that the current bill will provide needed support for academic supercomputing, it is unclear as to how this will directly support the pressing national need for very high performance commercial networks.

Title III: "National Information Infrastructure"

Twenty-five years ago, the President's Scientific Advisory Council recommended the establishment of federal "information centers" to support the conduct of science in the United States. In the succeeding 25 years, information centers have not emerged as a key element in the scientific information "system", although on-line database services have expanded enormously. Interestingly, the report containing these recommendations (*Science, Government, and Information,* 1963) made strong distinction between retrieval of information and retrieval of data or documents, specifically recognizing the need that scientists have for retrieval by content. The report envisioned centers staffed by scientists and other professionals who, it was hoped, would better serve the needs of the scientific community by providing analytical services as well as retrieval of raw data. In the succeeding two and one-half decades, on-line line databases have served the latter needs and serious scientific journalism (including semi-popular scientific newspapers and journals like *Scientific American* or *Discovery*) the former ; the concept of "information centers" as such has not developed despite the encouragement of the federal government. The development of "national knowledge banks" as recommended in this title of the Act should benefit from critical review of the experience of the past twenty-five years in the area of information sciences.

Title IV: "Software"

I fail to see the benefits to be expected from preferential support of artificial intelligence research relative to other aspects of software development as discussed above. Specifically, experience of the past few years has shown convincingly that the original optimism of the AI community was ill-founded. The exciting problems which they enthusiastically tackled have proven difficult and the solutions and methods much more limited in scope than had been hoped. Although I favor continued federal funding for support of AI and related areas of science and technology, I do not favor that it continue to be given preferential treatment within the computer science domain. I support the other provisions of this title, e.g. establishment of clearinghouses for distribution of re-usable code resources.

Title V: "Computer Systems"

I strongly support the provisions of this part of the Act, including the relaxation of overly restrictive export controls which inhibit the United States from effectively competing in the international marketplace and contributing positively to the balance of trade equation. I support as well the relaxation of rules which currently require that government subcontractors forfeit rights to proprietary software development tools used to develop advanced software systems.

Title VI: " Basic Research and Education"

Many of the developments I have described will be made in the area of software. As I noted, the best software is written by a small fraction of the total number of computer scientists. Better methods of training computing professionals are needed to improve this ratio. Also, significantly more research is needed in the areas of distributed and parallel computing software if the full power of these advanced computing architectures are to be realized in practical settings.

Senator GORE. Why do you not show your two to three-minute videotape now. I understand it demonstrates ways Apple is using its Cray supercomputer.

Dr. NAGEL. That is right, and I think it is self-describing. It does have an audio track, so if you just turn that up a bit.

[Videotape was shown.]

MODERATOR. Experiencing warpage and esthetic difficulties in production. In order to understand the current problems with the part, a wire frame model of the production part is produced. Then a five-element triangular mesh is applied to the wire frame for the beginning of the analysis.

The Macintosh is acting with a Cray XMP/48. As you can see, this level of computing power allows an engineer to obtain a level of detail which had traditionally been impractical for this type of analysis. Traditionally, accuracy in the analysis has been sacrificed in order to obtain timely results.

The raw computing power of the Cray XMP allows the engineer to explore the final details of the model without having to wait an excessive amount of time to obtain the results.

Dr. NAGEL. In this case what is being designed is a keyboard for a personal computer.

It is embedded somewhat in the film, but in this case what you see is the real-time computation of the Cray.

MODERATOR. The detail of the model can be more fully appreciated.

Dr. NAGEL. The display is actually being done on one of our Macintosh computers.

MODERATOR. The image that you are now seeing is a simulation of material flowing into the cavity to form the part. Notice where the flow fronts are meeting to form weld lines.

Color changes signify only change in time during the filling process. As you can see, large weld lines are forming in highly visible areas of the part.

Senator GORE. This is simulating the manufacturing of that?

Dr. NAGEL. That is correct, it is simulating the flow of plastic materials into the mold if it were made that way.

MODERATOR. Notice when the flow fronts are meeting to form weld lines. Color changes signify only change in time during the filling process.

As you can see, large weld lines are forming in highly visible areas of the part. In this particular production setup, 15 weld lines are formed. Due to the mesh density allowed by using the Cray XMP/48, we were able to run the analysis in approximately one hour and predict all 15 weld lines within approximately one-eighth of an inch of their actual production locations.

During filling, also notice that the part is not filling uniformly. One side of the part fills much sooner than the other side of the part, causing a large pressure imbalance in the cavity.

Since this part has been experiencing warpage problems, this imbalance should be improved, if possible.

Dr. NAGEL. Now, this is a case in which we did not catch it until after the fact.

I think that is fine.

Senator **Gore**. Well, presumably a company that did not have the resources to buy a Cray at the present price could gain similar competitive advantages by getting access to a Cray over a network.

At this point let me put into the record by reference an article from the Japan Economic Journal [1] about MITI launching a ten-year 50 billion yen program in computer-integrated manufacturing. They will use networks to make possible an intelligent manufacturing system with standardized factory operating systems so that different manufacturers' robots and computer information systems could be integrated.

As a first step, they are seeking a 1990 budget allocation of one billion yen for research and planning MIT plans to establish by next spring an organization with 100 Japanese producers and users as members who will use fiber optic cable to link factories and computer systems so that surplus manufacturing capacity throughout Japan could be instantly used to produce different items. And of course they have a plan to make it international as well.

Let me go to our last witness, Dr. Raj Reddy, the Director of the Robotics Institute at Carnegie-Mellon and President of the American Association for Artificial Intelligence.

Thank you for your patience, Dr. Reddy. Please proceed.

Senator **Pressler**. Mr. Chairman, could I just put some questions here, because I do want to listen as much as I can. I have been trying to deal with a couple of problems this afternoon.

But to the entire group, the question is the product of the Earth station, that if we build the EOS, when that information comes down, particularly in the area of artificial intelligence, and I know that that mimics the human intellectual processes, but are there ways that we can use that information?

How would that fit in with the development of artificial intelligence, some examples of it?

I am trying to carve out an area here and make a contribution. I am also trying to have a facility in my state play a key role in it, namely to receive these pictures.

But I attended a meeting in the Vice President's office recently and the Earth Station data was almost completely overlooked, some of the imaginative spinoffs that are possible. And that is not your subject, but I give that question to everybody in case I have to leave before the end. You can submit it in writing or orally.

STATEMENT OF DR. RAJ REDDY, DIRECTOR, THE ROBOTICS INSTITUTE, CARNEGIE-MELLON UNIVERSITY, AND PRESIDENT, AMERICAN ASSOCIATION FOR ARTIFICIAL INTELLIGENCE

Dr. **Reddy**. Senator, I used to be the chairman of the computer science committee of NASA a few years ago, and there are a number of potential spinoffs of that technology. In particular, the data that has been collected, as has been pointed out, have never been looked at significantly.

It turns out if we have a permanently manned station which is a way station to the Moon and Mars and exploration of the planets, then a number of things become possible here on Earth as a spinoff of this technology, which I would be happy to submit to you in writing afterward.

---

[1] See Japan Economic Journal, July 15, 1989.

Senator PRESSLER. Thank you very much.

Senator GORE. Let me invite all of our witnesses to provide responses for the record to Senator Pressler's inquiry.

Dr. Reddy, if you want to proceed with your statement, then we will—I will turn first to Senator Pressler for questions to the entire panel.

Dr. REDDY. Mr. Chairman, I want to thank you for this opportunity to be here and comment on S. 1067. As you know, the information industry is the single largest industry in this country today. It surpasses energy and automotive industries; it passed a few years ago. It is estimated to be between 10 to 25 percent of the GNP, depending on all the things that you count into it. It might even be 40 percent if you count all the Federal knowledge workers into the industry.

But the problem we are facing is we may be beginning to see the beginning of the end of the U.S. supremacy in this area. We have already lost our lead in the DRAM memory chips, optical disks, laser printers, laptop computers, and robotics.

We are beginning to lose the dominance in supercomputers, RISC processors, software engineering, expert systems, and distributed processing. And one very visible example of this is, if you were to have the NeXT people and SUN bring their latest work stations in front of you and open up and look at and see what is in there, 90 percent of all the high value added components today are produced in Japan or in the Pacific Rim countries.

My feeling is if personal computers and work stations become the commodity computers by 1995 or so and account for 75 percent of the market, then we will not have supremacy in this industry.

The decision to eliminate the export license requirements of certain classes of personal computers is one of the first steps to help make our industry competitive in the world markets.

Now, I would like to make a few comments on the advanced software initiative and the AI, as you asked in your letter. You asked for a short list of what could be done, where the investment might be made. I thought about the same issue and there are a few things I mention in my written testimony. I would like to comment on a few of those, not all of them.

In particular, the software repository idea is very important. The issue of having reusable, object-oriented software modules, not just any old software but reuseable modules, documented and available on-line, would be of immense importance.

It does not exist today, and in order to achieve it you need some research and some understanding and some standards that do not exist today.

Secondly, a complementary thing that would be very helpful is a handbook of software engineering practice.

Senator GORE. Of what?

Dr. REDDY. Handbook of software engineering practice, just like engineers, all engineers, have handbooks. We do not have such a thing.

As Bill Wulf and Jack Schwartz mentioned, software is an art, and we understand a large part of this art. It is about time we begin to codify this knowledge and make it available.

The third component is: it is not enough to have a repository or a handbook if people do not exist who understand where the reusable modules are and how to use them. So a major part of what we need to do is to train people, and this is something that is said over and over again in every testimony I have seen.

But I think there is something you could do very deliberately. If you asked every Federal contractor today and said, how many trained computer scientists with bachelor's, master's, and Ph.D.'s in software science, software engineering, knowledge engineering, and other related topics, you will find that there are a very small number.

Most of us have come into this discipline from other disciplines. I am probably one of the few people over fifty that have a Ph.D. in computer science. I happened to be hitting the wave about that time when the computer science departments came along.

But I think requiring trained people to be available to do all the software projects and giving preference to those aerospace industries, who have a lot of trained people, would do a lot to cause training to happen. Right now there are not many trained software engineering professionals.

An associated thing of that type that would be very useful is to provide for continuing education for existing software programmers and engineers. There is not such a program now.

The same gigabit network that you are talking about can be, with appropriate modifications, used for continuing education to train the people we need in this trillion dollar industry Dr. Wulf mentioned earlier that we will have to deal with by the turn of the century.

I would like to speak for a few minutes about second set of topics. We talked about the grand challenges of science. There are also grand challenges of society we can help to solve with the use of this technology.

For example, you may be aware that over 40,000 people die in this country in automotive accidents. In comparison, only a handful of people die in nuclear accidents today. And I think the technology, the computer technology, can be harnessed to eliminate over 80 percent of those fatal deaths by having accident avoidance features, what you might call an autopilot.

If you are drunk or fall asleep or not paying attention, you do not have to get into an accident. That technology is here. In fact, if you and Senator Pressler would like to come to Carnegie-Mellon, under the support of DARPA we have a Chevy van which navigates around obstacles and is able to do all those things, except it uses currently a $200,000 computer system.

None of us can afford such a car, but I believe, given the orders of magnitude improvements we are getting in computer technology, in less than ten years an autopilot option can be available on every car for under $1,000.

But somebody has to cause that to happen. Even otherwise, it will happen in less than 20 years. Already, the Mercedes Corporation in Germany, have a similar car. They have a car driving on the autobahns at very high speeds, except their solution is not as advanced, we believe, as we have.

The second, similar grand challenge for the society that the computer technology can help to solve, is the area of the translating telephone. One of the problems we have is that we Americans are the worst in being able to communicate in other languages. Everybody else masters languages and as a result they are able to compete effectively in a global economy.

We cannot continue to do what we have done in the past. We have to become more literate in Foreign Languages. And again, computer technology can help in this area.

The Japanese already have a first phase project of $120 million on building a translating telephone, and they expect to spend over $850 million in the second and the third phases.

All the technology is not here. It will take a lot of research and development. It may be about 25 years to do it, but it can be done.

The last one, which is very important, you mentioned manufacturing and visualization. I believe we have a specific example. In the past computer visualization and computer simulation was only used for training pilots, for example, to give them a realistic rendering of what they would see when they are flying a plane. But it needed a $10 million simulator that was built by Evans & Sutherland.

We built a factory simulator so that every worker could be trained on the machines he has to operate using the current generation visualization systems. Again, if you have an opportunity to come to Carnegie-Mellon we would be happy to show it to you.

This was estimated to save Westinghouse Corporation over $5 million every year. Because of the turnover of their employees, they were making costly mistakes on the factory floor, which they were able to eliminate by this simulation and training.

Factory simulation broadly can be significantly enhanced using the lower cost visualization technology that is now possible.

I would like to conclude with a few comments. Much of the development proposed in S. 1067 is not only important, but may be essential for maintaining the U.S. technological lead. Indeed, it may be too little, too late. The projected investments of a few hundred million dollars a year may be too little if the information industry continues in its current slump.

The industry is at present concerned with diminishing stockholder value, and it has never recovered from the 1987 crash-and they are all in the cellar. And it is preoccupied with takeovers. In such an environment, strategic investments for R&D are the first to go.

Other countries, such as Germany and Japan, with focused industrial policies on strategic investments do not depend solely on private industry initiatives for long-term R&D. Unless we are willing to accept the near certainty of becoming a second-rate industrial power in the twenty-first century, we can no longer afford to do that either.

One possibility is an R&D set-aside of five to ten percent of current expenditures in all the Federal government information industry expenditures. It has been estimated DOD alone spends $30 billion a year on software acquisition and maintenance. If even five percent of that were put into R&D to see how that could be done better, we would have more money per year than Senate Bill 1067 is proposing for five years.

Similarly, overall it is estimated the government spends over $100 billion in this area. But most of it goes into various acquisitions, very little to technology insertion to see how the emerging technologies can do these things better and cheaper.

One of the things we find is, again we were shocked to see, because of the procurement regulations and various other requirements, a lot of the technology that is being used currently in the government is 10 to 20 years out of date. And somehow we need to find a way of overcoming that problem, and one of the ways is in fact taking some of that money and asking it to be spent on technology insertion.

Mr. Chairman, this calls for an aggressive industrial policy on the part of this country, and our security, our industrial base, and our destiny depend on such an action. As President Bush recently said: "Destiny is not a matter of chance." It is a matter of choice. And I hope we will all pursue that.

Thank you.

[The statement follows:]

Statement of Dr. Raj Reddy

University Professor and Director of the Robotics Institute
and
President of the American Association for Artificial Intelligence

Good afternoon Mr. Chairman . I thank you for the opportunity to testify before your subcommittee on Senate Bill 1067 and, in particular, Title IV of the bill on Advanced Computer Software. Mr. Chairman, your leadership and vision in this area of vital interest to the security and economic well-being of the nation is highly appreciated. As a University Professor and Director of the Robotics Institute at Carnegie Mellon University and as the President of the American Association for Artificial Intelligence, I have had first hand experience with many of the concerns and goals of the Senate Bill 1087. My testimony today reflects the views acquired over thirty years of professional experience.

A Crisis in the Making

Mr. Chairman, you know that the United States has been and continues to be preeminent in the information industry. This industry contributes over 500 billion dollars to the GNP and has surpassed the energy and automotive industries in its contribution to the GNP in recent years. It employs more than five million people. It is estimated that the industry paid well over twenty billion dollars in taxes and close to 10% of individual income tax dollars come from Information Industry employees. In fact, it is an industry which we can ill afford to ignore and leave to chance.

The United States is still in a leadership position in the information industry. However, we may be seeing the "beginning of the end" of the U.S. supremacy . We have already lost our lead in DRAM memory chips, optical disks, laser printers, lap-top computers and robotics. We are beginning to lose dominance in super computers, RISC processors, software engineering, expert systems and distributed processing.

To understand the seriousness of the problem, all we have to do is to look at the best recent computers from NeXT and Sun Microsystems. If you open the box and take a look at the contents, you will find that over 90% of all high-value components and subsystems are made in Japan. It is only a matter of a few years before Pacific Rim Countries will dominate the commodity computer market, i.e., the personal computers and the workstations which will account for over 70% of the hardware sold by 1995. The decision to eliminate the export license requirements for certain classes of personal computers will help to make us more competetive, but it is only a first step.

Advanced Computer Software

It is estimated that software contributes 70% to 80% of the $500 billion GNP number mentioned earlier. More specifically, it is estimated that DoD spends over 30 billion dollars annually on software, 80% of which is said to be for maintenance of existing software. A recent article in *Businessweek* cites that almost every high tech weapons program is significantly behind schedule and over budget; software appears to be the major culprit.

Research proposed in Title IV of Senate Bill 1067 will help significantly to alleviate these problems. While hardware speed has been increasing at over *3% compound growth rate every month*, leading to a thousand-fold improvement over the last twenty years, software productivity has been creeping along at only about *4% to 6% compound growth rate* ***every year.*** Meanwhile demand for software in certain sectors has been increasing at over 20% per year, leading to a *software gap*. The cost of software has become the dominent cost in the use of computers, rising from 20% to 80% of the total cost of a computer installation.

Recent initiatives at DARPA on languages and systems for Rapid Prototyping of software are a welcome change in this area. If significant additional funds are available, a number of steps can be taken to improve the productivity and quality of software:

- Creation of a repository of object oriented reusable software modules for every major application domain.

- Creation of a handbook of software engineering practice (preferably on-line) for all application areas.

- Develop centers of excellence at all major universities for training of graduates and retraining of existing software practitioners in advanced software tools and techniques.

- Make it a prerequisite that all federal contracts involving software development use trained graduates (or retrained professionals) from accredited universities in computer science, software engineering, knowledge engineering and other approved majors.

- Create demonstration "Shadow Projects" on selected federal contracts to validate the claims of higher productivity of properly trained software engineers.

- Support the purchase of commercial software tools by researchers in application areas such as chemistry, physics and material science.

Opportunities and Benefits in Artificial Intelligence.

Your invitation also asked for comments on research and potential benefits of Artificial Intelligence. Artificial Intelligence is one of the areas that is critically dependent on all aspects of High Performance Computing Technology: hardware, software and networking. If we are successful in High Performance Computing, especially advanced computer software, it will become possible to realize some of the grand challenges of Artificial Intelligence of immense benefit to society. The following is a list that highlights some of the national problems that appear solvable through the use of Artificial Intelligence technology based on High Performance Computing.

- *Accident Avoiding Car*

  In the United States, over 40,000 people die annually in automobile accidents . It appears that a new generation automobile equipped with an intelligent cruise control using sonar, laser, and vision sensors could eliminate 80% to 90% of fatal accidents and cost less than 10% of the total cost of the automobile. Such a device would require research in vision, sensor fusion, obstacle detection and avoidance, low cost/high speed (over a billion operations per second) digital signal processor chips, and the associated software and algorithm design.

  Solution to the problems of obstacle detection and avoidance have been developed, in recent years, under the sponsorship of the Defense Advance Research Projects Agency. However, the cost of computers for each car using the current technology would exceed $100,000. High Performance Computing would accelerate the availability of low cost super computers and associated advanced software modules.

- *The Translating Telephone*

  Given our trade deficit, we badly need to become a global competitor. In a global economy, we Americans are at a significant disadvantage. Our ability to read or write a business letter in Japanese, German or Russian is almost non-existent. Computer assisted reading, writing, and conversation in multiple foreign languages may be possible within the next decade. However, it will be used only if High Performance Computing Technology Initiative leads to low cost super computers and advanced software for multilingual communication.

  Japan recently initiated a seven year $120 million project as the first phase toward developing a phone system in which a Japanese businessman can converse with, say, an American businessman in real time. This requires solutions to a number of currently unsolved problems: a speech recognition system capable of recognizing a large (possibly unlimited) vocabulary and spontaneous, unrehearsed, continuous speech; a natural sounding speech synthesis preserving speaker characteristics; and a natural language translation system capable of dealing with ambiguity, non-grammaticality, and incomplete phrases. In the United States we have no focussed program in this area of research.

- *Intelligent Learning Environments*

With a successful launching of a High Performance Computing Technology initiative, it will be possible before the end of the century, for each person to have access to a super computer on a desk top. What would an average person do with a super computer on a desk? Many experts seem to agree that there is a crisis in the quality of education at the high school and undergraduate levels. Furthermore, it is no longer adequate to attend school and college for twelve to twenty years and plan to coast for the rest of life. In a rapidly changing technological world, it will be necessary to have facilities for life-long education. It is in this area that a desk top super computer will provide large dividends by making it possible to have personalized learning and education.

High Performance Computing and Advanced Software Tools can significantly transform the processes of learning and education. The most exciting prospect for me is the possibility of creating the electronic equivalent of laboratories and libraries.

In an electronic laboratory, a student can conduct or recreate physics and chemistry experiments, observe the behavior, modify the conditions (even beyond current physical limits) and gain a deep understanding of the processes that can never be attained from books. The technology will have a profound impact on poorly educated enlisted personnel within DoD who are required to handle increasingly complex hitech weapons and systems. It will permit quality universal education for students who have no ready access to test tubes, Bunsen burners, chemicals, and instruments that students in a rich environment take for granted.

The other exciting possibility is the "digital library." It appears that advances in computers and communications will make it possible for anyone, anywhere in the world, to access and read any book, report, magazine, or newspaper, in any language at a cost of less than a cup of coffee.

- *Virtual Symbolic Factories*

Another major use of High Performance Computing Technology will be in Knowledge Based Simulation in manufacturing. KBS is a technique which can have significant impact on the process of invention, refinement, and adaptation of products to varying customer needs. Unlike wind tunnel simulation which uses finite element techniques, KBS uses knowledge about the task such as facts, rules, and constraints to create a symbolic simulation of the activity. Thus, a symbolic "factory" would contain descriptions of the machines, capacity and throughput, layout, tools, fixtures, raw materials and processes.

KBS techniques are being used in factory simulation tasks such as production planning, resource planning, layout planning, dynamic scheduling, and operator training. Use of these techniques is not yet widespread because of the need for an interdisciplinary team of experts required for codifying the relevant knowledge into an expert system. As this technology spreads it can be expected to have serious impact on productivity, quality, cost and timeliness which may result in widening the gap and disparity between the efficient producers such as Japan and inefficient producers. It is essential that the U.S. use such tools to become a world class manufacturer, to compete in the global market and to eliminate the trade deficit.

What Can We Do?

Much of the research proposed in Senate Bill 1067 is not only important but may be essential for maintaining the U.S. technological lead. *Indeed, it may be too little, too late!* The projected investments of 50 to 150 million dollars a year may be too little if the information industry continues its current slump. The industry is at present concerned with diminishing stockholder value (most computer stocks have never recovered from post crash lows of 1987), and preoccupied with the threat of takeovers. In such an environment strategic investments such as Research and Development are the first to go. Other countries with focused policies in strategic industries do not depend solely on private industry initiatives for long-term R&D. Unless we are willing to accept the near certainty of becoming a second-rate industrial power in the twenty first century, we can no longer afford to do so either.

In an era of constant or diminishing budgets it may appear to be difficult to justify significant increases in Research and Development in Information Technology. But we must not forget the following facts. Currently the Information Industry accounts for over 10% of the GNP and an equivalent contribution to U.S. taxes. In addition, it is estimated that the U.S. Government spends over 100 billion dollars on information technology related expenditures annually. Furthermore, the cost of computing has been decreasing at an exponential rate. It is said that if the automotive industry were to improve at the same rate, you could buy a Rolls Royce for under ten dollars, which would deliver 100,000 miles per gallon, and go at a speed of a million miles per hour. Unfortunately, much of this performance has not found its way into federal computer use. The problems are: complex procurement procedures in computer acquisition cycle, lack of trained personnel, and technical complexities associated with technology insertion into an ongoing operation.

One possibility is an *"R&D Set-Aside"* of 5% to 10% of current expenditures to explore how the latest technologies can do the same job at a much lower cost. A five to ten billion dollars a year investment into R&D in the areas discussed in Senate Bill 1067 may be necessary to improve efficiency and to reduce the cost of federal computing expenses, and to reverse the trend and regain momentum in the vital Information Industry.

Mr. Chairman, this calls for an aggressive industrial policy on the part of the government in an area of strategic interest to the country. Our security, our industrial base, and our destiny depend on such an action. As President Bush recently said, "Destiny is not a matter of Chance. It is a matter of Choice."

Thank you.

Senator GORE. Thank you very much.

Senator Pressler.

Senator PRESSLER. I read with great interest a page of testimony—I believe it is Dr. Clark's testimony, the second page of it—about the national and education network. And I am going to ask, first of all, now the present mechanism connecting computers is—help me out. How many gigabits can it translate per second or minute compared with the proposed fiber optic network?

Dr. CLARK. The present rate, the most common interconnect scheme used is the ETHERNET technology, and that is a ten megabit per second technology. So there is a factor of 300 in the proposed three gigabit network versus that. And of course, fiber optics is capable of factors of a thousand and several thousand more than the kind of technology that is the predominant technology that we use in connecting all of our computer systems today, the vast majority of them I should say.

Senator GORE. Could I interject a point there, because I think the point you are getting at is a very important one. You are talking about ten megabit or ten million bit a second over relatively short distances, right?

Dr. CLARK. That is also correct, yes.

Senator GORE. Over the longer distances of the kind we hope to span in this network, you are really talking about one and a half megabits?

Dr. CLARK. Yes, that is right. It is about an order of magnitude less over long haul.

Senator GORE. So 1.5 million bits a second versus 3 billion bits per second.

Dr. CLARK. We are talking between a thousand and 10,000 times speed-up.

Senator PRESSLER. Now, the visualization supercomputers that are available today that are hooked together, how are they hooked together?

Dr. CLARK. There again, the predominant means of connecting them together is over the ETHERNET. There just has not been a standard evolved.

Senator GORE. So as a result, some of the most valuable ways of presenting information cannot be used by a researcher using a supercomputer in a different geographic location.

A real-time conversation between supercomputers in visual imagery really cannot take place today.

Dr. CLARK. There exists no current method for that.

If I might interject also, the side benefits of focusing on this fiber optics technology will be, I believe, that there will simply be a much higher bandwidth connection between all computers, and therefore a collection of computers could collaborate in solving a very large problem.

Namely, the massively parallel problem could become, frankly aside from software, which I do not want to diminish, but the massively parallel problem could become solved by simply having a high bandwidth network available with that technology.

Senator GORE. If my colleague will indulge me just one more minute, let us say for example that a researcher at Carnegie-Mellon wanted to use a massive parallel supercomputer to look all over the surface of the Earth for volcanic activity. As you know, many volcanic eruptions are never reported, but they are visible on LANDSAT photographs, 90-plus percent of which have never even been looked at.

Let us say that a researcher in Pittsburgh wanted to connect to the data base in South Dakota and use the capacity of a massively parallel advanced computer to sort through tens of thousands of images and identify volcanic activity all over the surface of the Earth. With today's networks that would be impossible.

But with a gigabit network, it would be possible, once the software was created, for researchers in Pittsburgh or in Oak Ridge or anywhere in the country to gain access to the data base in South Dakota and make a tremendous use of it, thus creating jobs for the constituents of the Senator from South Dakota.

Senator PRESSLER. We hope so, but also serving our country well.

Let me say on that point that I am preparing a letter to Vice President Quayle, who is heading up an effort in the administration, and also I am going to make a Senate floor speech. And I have given this question to the panel.

I tend to believe that the space program from my point of view should be focused on Earth study at this point, and that is the only way I can sell it to my constituents, very frankly, and that is what I happen to believe very strongly. But how we relate all that information to this vast fiber optic network connecting up the supercomputers, that is a very important thing.

But also taking a look, as in your testimony, Mr. Reddy, at the use of this in artificial intelligence. Most people just think, well, we will look at the pictures and see what is there. But I think we have to think a step beyond that, and you have done some good models here on artificial intelligence in other areas.

It might be hard to think of examples, but I think examples of that are what we could best sell the space program or the Earth station, whether it is the LANDSAT pictures or something different, but it is in the same category—but that is something that we need to really think about.

Dr. REDDY. Senator, there are a large number of components that go into making that data useful. First is to have a backbone network.

Senator PRESSLER. First is to have what?

Dr. REDDY. To have a backbone network of the kind we are talking about.

Second is a data repository which is collecting this data and indexing it and storing it away in various forms, such as the one that I believe currently exists.

Senator PRESSLER. Good, music to my ears, at the EROS Data Center at Sioux Falls, South Dakota.

Dr. REDDY. Exactly.

The third, the third component, is to then be able to process that data using techniques of image processing and image understanding of the kind that have been developed under sponsorship of DARPA and NSF over the last 20 years, with specific objectives in mind.

If they happen to be objectives of environment, objectives of pollution, or objectives of urban or agricultural growth, all of that can be done. But that requires discipline-oriented studies.

So somehow, once we say we want a problem to be solved, then that in turn defines what research has to be done.

Senator PRESSLER. Well, that is true. Like for example, on page 2 of your testimony, I believe, or page 1, you talk about how big the information industry is and how many people it employs. But some people would ask, is that necessarily productive? You would have to say, not unless somebody is doing some models on some artificial intelligence, not unless it is being used.

A lot of people then say, well, we are becoming a service and an information industry and country and we may collapse by it, it may collapse by its own weight; is it productive? Well, of course it is productive.

But I think harder than getting, gathering the information and distributing it, is going to be to figure out how to use it profitably. And that I suppose makes the artificial intelligence area the high C's and D's of the opera, so to speak.

Dr. REDDY. Yes and no. I think I agree with what Jim Clark was saying earlier. There are all these components—networking, software, high performance computing, advanced computer technologies, which include artificial intelligence techniques, and application areas—that together form the glue.

No one of these can stand on its own. Without the high performance computing and the networking and the advanced software, we could not do the AI systems I talked about. In particular, if anybody tells us, attempts to tell you that the computer information industry is mainly a service industry, they are completely wrong.

Every automotive manufacturer produces 80 pieces of paper with every car they produce, and five percent of the design engineering cost accounts for 70 percent of the total cost of an automobile. So if they do stupid things, then they impact the entire 70 percent of the cost of the car.

And it turns out information technology is so pervasive, we could not simply label it and put it in the bucket of service industry. Across the board, it affects every facet of what we do, in particular manufacturing and design.

Senator PRESSLER. The one other area of curiosity about your testimony, you talk about translating machines and perhaps you could talk on a telephone on a fiber optic cable with someone in a different language. How far off are we? As you say, we do not have all the different languages and the different accents and so forth. It is probably a way off.

Dr. REDDY. Yes. Again, under the sponsorship of DARPA, we demonstrated a system at Carnegie-Mellon where any speaker can come and speak. It is a speaker-independent, connected speech system with about a thousand word vocabulary, going from there to multiple languages.

The Japanese are working on a similar problem and we hope to demonstrate a simple business dialogue in a narrow domain within the next two, three years between ATR, which is in Japan (it is a subsidiary of NTT) and CMU.

That does not mean that problem is solved. We are probably 20, 30 years away before an arbitrary person can talk about an arbitrary topic. However, we do not have to wait that long.

What we are saying is there are a number of areas—if you want to write a business letter or understand a business letter written in Japanese or German, we currently cannot do it. And that significantly limits our ability to do business in a global economy, and we are going to be significantly falling behind because of it.

I believe all the technologies are coming along and useful solutions can be built within the next two to five years, and they will get into the marketplace in five to ten years, essentially the same kind of model that Jim Clark was talking about.

Current work in progress will make that happen anyway, but the question is can we do it faster and be dominant in the world market?

Senator **Pressler**. Well, I appreciate your comments on the artificial intelligence area as it relates to some of the matters I have raised, and I look forward to anything additional you can submit.

I apologize to the chairman and I thank him for letting me go first here. I have to run to a meeting regarding funding one of our Indian tribes who are having some problems, and they want this Senator to get them some money.

Senator **Gore**. Thank you very much, Senator Pressler, and thank you for co-sponsoring the bill. This is a partnership project here.

Dr. Clark and Dr. Poduska, your companies both build computer hardware that is specially designed for visualization. How are your graphics supercomputers different from regular supercomputers?

Dr. **Poduska**. Well, to start with they are designed intentionally to be a balance, an economic balance of computation, both scalar and floating point, graphics capability, and I/O capability. The intention of the design, that is the design point itself, is to put something that is economically feasible in the office with a scientist or an engineer, that makes a reasonable amount of noise and burns a reasonable amount of power, but provides them with computational capability for generating the models that they look at—the isosurface planes that I was talking about, pseudocoloring, and that kind of stuff, and then the graphical capability to actually look at that, to cause it to be rendered as a visible, viewable kind of image.

Now, the technology marches on at an uneven clip between these two different things, and so the task of finding an economic balance between these things, these different elements of the thing, of the graphic supercomputer, is an ongoing task and one that is subject to debate.

You would find differences in the machines that are available from Silicon Graphics and Ardent and Stellar and around. But I think the basic notion behind them is the same: balance what you see, the graphics and computation, specifically with the thought in mind of making the scientist or engineer able to see the results of the computation.

Senator **Gore**. Let me come back to that. But I want to hear your response, Dr. Clark.

Dr. **Clark**. I could just amplify on that, the concept that basically at Silicon Graphics we recognized up front the need to have visualization balanced with computing. And so it is a balanced system. The objective is to treat visualization as just as important as the computation itself, because it is only through that means that you can comprehend the computations.

The other difference is the notion of putting together enough microprocessors in a single device to start to achieve supercomputing speeds at an economic price.

Senator GORE. Let us look at the rate of progress in different parts of this enterprise. Dr. Winkler, you talked about the relationship between the capacity of the eye-brain system to absorb information and the capacity of the computer to generate it and present it.

According to your testimony, our computation and presentation capabilities have already gotten past that threshold.

Dr. WINKLER. Yes.

Senator GORE. Okay. Now, if you look at processing capacity, visualization software, communications and networking, and data storage, how would you rate those five in terms of progress made? Processing has gone the farthest down the road, correct?

Dr. WINKLER. Yes.

Senator GORE. What is next among these four items?

Dr. WINKLER. I would say, what is not allowing a balanced environment for the type of work I have showed you here is storage, data storage.

Senator GORE. Data storage.

Dr. WINKLER. Data storage is the main one, and communications. The network you are proposing in this bill would make all the difference, as far as I can tell, on a short time scale. But it is storage and communications in my mind.

Senator GORE. Storage and communication.

Dr. WINKLER. The network itself.

Senator GORE. All right. The network proposed in the bill would improve communication between computers, but your data storage problem would still be there?

Dr. WINKLER. That is correct, and it is there for many other reasons. Certain types of computations simply don't get done although we have the algorithm, and even the visualization, but we just cannot store the intermediate results.

Senator GORE. Anybody else want to respond to the same question?

Dr. CLARK. I just might add that we lead the Japanese in networking and we lead the Japanese, or at least we still have a toehold, in the mass storage market. There are domestic suppliers of mass storage devices and, by contrast, there are no domestic suppliers, of any significance at least, in the memory, the dynamic memory area.

The vast majority of the components we buy from the Japanese, but in the area of storage and networking we are ahead of the Japanese.

Dr. REDDY. Even there we are beginning to lose. For example, the most recent optical disk in the NeXT computer is made by Canon. And we are neck and neck in the storage area and they are beginning to take the lead, and we are losing our supremacy in one after the other after the other, and there does not seem to be an end to it.

Dr. NAGEL. It seems like every part of the industry in which process is important is an area in which slow domination is inevitable or has been inevitable in the past. Areas in which concepts are paramount are ones in which we still—as you say, you have a toehold or in some cases a lead.

Anything where process, improvements in process technology can lead to domination, that has happened.

Senator GORE. If we pass this bill, which has not happened yet, would a useful next step be to look at a priority effort in data storage?

Dr. NAGEL. I would actually like to follow up on the comment that Jim Clark made earlier, and that is somehow to encourage the expansion of this national network idea, so that individuals in their homes can take advantage of it as well.

Senator GORE. Right. Well, that is part.

Dr. NAGEL. And I understand that is a different problem and it is part of different legislation. But I think it is one that the Federal government could somehow encourage the kinds of investments in the infrastructure in the telecommunications industry, for example, that would greatly stimulate that, even though there are local regulatory problems that obviously have to be overcome.

But I think we also think that is a real important area to begin in.

Senator GORE. I agree with that. I have every intention of doing exactly that and trying to see that the Senate and the country does that. I will certainly be involved in that effort.

But let me just ask for a minute about data storage. Let us just say as a thought experiment that you wanted to digitize the entire Library of Congress and you were able to digitize text in a mechanized way that did not require constant human input.

How would you store it?

Dr. WINKLER. I think the difference—the point to make there is that you have text and you have visual information. And as you pointed out in the introduction to this session, a picture is worth a thousand or a million words, depending how you count it.

So whenever you include images in the storage system, your storage requirement goes up dramatically.

Senator GORE. But you can store an awful lot of still pictures on most storage media.

Dr. WINKLER. Yes. What I am concerned about is also animation.

Senator GORE. Yes, if you have a lot of animation then you are going to eat it up very, very quickly.

Dr. WINKLER. I think with current storage technology, certainly off-line, we can have an archival system. If you want to have really access to the data, then the question becomes one of how fast you can access them on top of whether you can store them. So it is really a combination of speed and capacity we need.

I do not think that we are really equipped to access all the information which our society has collected.

Dr. PODUSKA. I think, sir, that I am perhaps a little more optimistic about these things than perhaps some of my colleagues. I think that if you look back over the last 30 years, you find that the cost-performance numbers of storage have increased even faster than the cost-performance numbers of silicon.

That was a revelation to me when I found it out, and that is true even of an old technology of magnetics, as has been our storage mechanism up until quite recently.

I think that when we get to the issues of storing images around, especially static images as your question suggested from the Library of Congress, there are lots of compression techniques, they are getting better and better all the time. I would be reasonably optimistic about the ability to put that in some sort of, for want of a better word, some sort of compact disk kind of storage mechanism.

I do not mean that precisely, but I mean some sort of laser-written, read many, written once, the WORM mechanisms as they are called around.

I think for the storage of data that comes in off of high-speed links, say satellite data, and need to be recorded, analyzed, perhaps corrected on the fly, that is a horse of a different color. That is an issue of being able to get fast access as well as rapid data transfer of information, and also an ability to read/write it probably, at least in the collection phase. And that is a much more difficult kind of problem.

Reflecting back to a question you asked a little bit earlier about the relative importance of the different components of things, in one sense it is kind of hard to separate those. I mean, it is kind of like which is more important, the heart or the lung. I mean, you need them all.

Senator GORE. Right.

Dr. PODUSKA. On the other hand, if you look where progress has been made, I think that here again I may be a little more optimistic than some of my colleagues. I view the United States relative to some of our Asian and European colleagues as being in a pretty good position on a number of things—processors, the graphics, technologies in general, software, especially individualized software, communications mechanisms, and at least until quite recently storage mechanisms.

I would not think of losing one inning as having lost the game, either, in that particular game.

Senator GORE. Dr. Clark?

Dr. CLARK. I would just re-amplify what Bill said in response specifically to your question, how would you store the Library of Congress. My instincts are that it would be some sort of laser disk technology. There, as Raj Reddy points out, unfortunately that is a derivative of compact disk audio technology, which has been pioneered by the Japanese and the Dutch.

But I believe that would be the way that it would be stored, and it would not take that many CD's to store it.

Senator GORE. All right. Unfortunately, we have a vote on the floor. That is what that last signal meant. It has been a long day.

The subcommittee has learned a great deal, from your presentations here today. I really appreciate all of you taking the time to be here for a long afternoon, but a very interesting and fascinating afternoon.

I want to thank Dr. Schwartz and Dr. Wulf again also. And to every member of this panel, thank you for helping us in this effort.

We are going to continue the work of trying to better understand the challenges we face here and hopefully pass S. 1067. And thank you very much again.

This hearing is adjourned.

[Whereupon at 4:58 p.m., the subcommittee was adjourned.]

NATIONAL HIGH-PERFORMANCE COMPUTER TECHNOLOGY ACT OF 1989

FRIDAY, SEPTEMBER 15, 1989

U.S. Senate,
Committee on Commerce, Science, and Transportation,
Subcommittee on Science, Technology, and Space,
Washington, DC.

The subcommittee met, pursuant to notice, at 9:35 a.m., in room SR-253, Russell Senate Office Building, Hon. Albert Gore, Jr. (chairman of the subcommittee) presiding.

OPENING STATEMENT BY SENATOR GORE

Senator Gore. The hearing will come to order.

I would like to begin by thanking everyone for their attendance here today. I appreciate it; the subcommittee appreciates it. We are looking forward to the exchange of information and dialogue that will take place here today as we examine the latest developments in the computer revolution that is fundamentally changing the way we deal with information.

I have a lengthy opening statement, but I do want to say just a few words before we get to our first panel of witnesses.

I am excited about the prospect of passing legislation this Congress to move the United States rapidly forward into the information age. We need to build a national network of information superhighways. We need the capacity to transfer massive volumes of data rapidly between different parts of the United States.

Just as the interstate highway system improved our ability to compete around the world, so a system of interstate highways for information will enable the United States to compete more effectively in the communications and computer fields as well in every other field.

Whatever line of scientific research or corporate research is being pursued, the ability to communicate the products of that ongoing research with others in the same field around this country will produce a kind of synergy or mutually reinforcing advantage as people can work together more efficiently.

Right now, though, we have a kind of chicken and egg problem, in that the network we need does not do this because we do not yet have wide use of the machines that need advanced networks. Yet, one of the principal reasons we do not have wider use of those machines is that we have no means of linking them together.

Eventually, the marketplace will enable this network to be financed by the private sector, but we need to get it over the hump, so to speak. We need to get it past the initial resistance, the inertia that is out there, and put it in place. Virtually everyone who has studied this idea believes that it is an exciting one that ought to be recognized as one of the Nation's top priorities.

In that regard I was delighted last week when Dr. Alan Bromley, the President's new science adviser, released the Federal High Performance Computing Program, a five-year plan developed by the Office of Science and Technology Policy. The plan is, of course, very similar to S. 1067, which we are going to be looking at in more detail here today, and calls for $1.92 billion over the next five years for advanced computing.

It is true, of course, that the President and OMB have not yet agreed to request funding for the program, but their statement of support for the idea is an important milestone. The science agencies are pushing hard to get funds for the program included in their fiscal year 1991 budget request.

I have had private conversations with physicists, astronomers, chemists, climatologists and others in a variety of different scientific fields, all of whom say this is what we most need.

Taking climatology as one example, we have been working on the problems of global climate change in the Congress. The different teams of researchers attempting to provide us better answers to the questions being asked by the American people now have to physically download their computer models on to magnetic tapes, take them on airplanes, travel some distance to visit with another team, and then upload the models.

When they finish collaborating, then they have to go all the way back to their own laboratories. They ought to be able to communicate on a minute-by-minute basis and mutually reinforce the fine research being done in a variety of different areas on global climate change. That is only one of many, many examples.

Japan recently announced a very ambitious plan to build a computer network that will link various manufacturing centers in Japan. The supercomputer manufacturers here in the United States say they engender more enthusiasm for their products among foreign manufacturers than among our own domestic manufacturers. That has implications for America's ability to be competitive because when the full potential of these new advanced computers begins to be realized they will make all parts of the manufacturing process far more efficient and effective.

In the field of education which we are going to be talking about a lot here today, there is also a tremendous opportunity. I have a vision of a schoolchild in Tennessee going home after school in the evening and settling down after dinner before a little machine that will look not much different from a Nintendo television game machine today. The child would have at his or her fingertips the information contained in the Library of Congress with the capacity to browse through that information at his or her own pace, directed by his or her own curiosity, asking questions and receiving answers, and maintaining a very high level of attention and intellectual curiosity. The relationship between the student and the information will be fundamentally different than that which exists when that student is trying to prop his or her eyelids up at night, slogging through a relatively dull textual presentation and attempting to glean the important bits of information, waiting for some new paragraph to once again reawaken the curiosity that might have been present at the beginning of the exercise but was deadened and dulled by the sheer tedium of plowing through the information as it is now so frequently presented to students.

I have had an opportunity to take part in demonstrations of the new technology for interactive computer presentations. It is really very exciting. Once you explore a subject using that technology, you immediately want to see that students all over this country have access to it and are able to use it.

Here again, the single best way to make that opportunity available to schoolchildren throughout this country will be computer networks capable of transmitting vast quantities of data. Where is the data going to come from?

Another part of the legislation that we will be talking about here today deals with something called digital libraries. Books can be stored digitally on magnetic tape, compact disks and computers and, using a computer, library patrons can access information from the books in the library's collection. Hundreds of thousands of books can be on-line, available at the touch of a button.

The technology already exists. Indeed, today you can buy a single compact disk containing the entire Bible, an unabridged dictionary, the complete works of Shakespeare and a half a dozen other reference books. Not only is it extremely compact and portable, but you can use a computer to search the disk very quickly and find a particular phrase or concept or reference much more rapidly than you could by hand.

Later in this hearing we will hear about commercial products offering hundreds of magazines on a single compact disk. Of course, a digital library would not really have walls in the traditional sense, because with computer networks people around this country could gain access to the digital information on-line at the touch of a button.

We are already seeing the beginning of a system of national digital libraries. A little bit later in the hearing we will hear from Dr. Jim Billington, the Librarian of Congress. The Library of Congress already maintains a national computerized card catalog which thousands of libraries nationwide access over computer networks.

In the same way, the Library of Medicine distributes medical information from its databases to doctors around the country using its Medline system.

The U.S. Geological Survey is developing the technology needed for a national digital map library. One of our witnesses today will provide a demonstration of this technology.

Let us be clear about one thing. Not only will this means of storing and distributing information make it easier and faster to gain access to it, but it will also make it practical for the first time for human beings to deal with the vast mountains of information that is now stored and yet never looked at.

I had a brief conversation with Mr. Billington last night. We were discussing this. And I have talked with my colleague, the Ranking Republican Member, Senator Pressler, about it.

Just as one example, consider the Landsat photographic images of the earth. Ninety-five percent of those pictures have never been seen by human eyes. The pictures have been taken; they have been stored; they exist; but they have never fired a single neuron in a single human brain.

It exists not as information, but as a kind of "exformation" outside the awareness of human beings, outside of our ability to deal with it. But, with this technology, we would not have to slog through every one of those pictures to find what we want. With computer processing we could instantly pick out the features that are of interest.

We could, with the help of expert systems, immediately catalog or quickly catalog geological features that were similar to one another all over the world, and vastly increase the understanding that geologists and geographers are now giving us of how the Earth's system works.

It has only been a couple of decades since scientists became convinced of continental drift, for example. What other discoveries of monumental importance are out there waiting to be made?

Well, they are there in the information.

To use another example, the ozone hole was discovered in the stratospheric ozone layer above Antarctica due to an almost serendipitous discovery by a British scientist which provoked American scientists to go back through satellite data that had already been collected. And, lo and behold, the data had been there for years. There was just such a large volume of it that nobody knew exactly what patterns to search for and exactly how to deal with the information, how to go through the information and learn from the information.

Well, rather than just storing information and never looking at it, we now will have the ability to deal with the vast quantities that are doing us no good at all, because we are not looking at them.

The research community is already working on solutions for all of the remaining technical problems that are involved here. I think S. 1067 provides the correct blueprint for moving forward rapidly to realize this tremendous opportunity.

Yogi Berra once said what we have here is an unsurmountable opportunity. I do not think this opportunity falls in that category. The Federal Government spent $63 billion last year on research and development. A tremendous quantity of the information produced by that $63 billion was data that was stored but never really examined.

By spending a tiny amount to give us the ability to look at the information, understand the information, and share it among different scientific and corporate groups around this country, we will get a much larger return on the investment in research and development, not only in the government expenditures, but in the private sector, as well.

As I said, I will put the rest of my statement in the record.

[The statement follows:]

OPENING STATEMENT BY SENATOR GORE

Today the Science Subcommittee will examine the latest developments in the Computer Revolution that is fundamentally changing the way we deal with information.

For thousands of years, information has been stored in one of two ways -- on paper or in people's heads.

But now we have a third ways, computers that can mimic the human brain. They can store billions of bits of information, sort through it all in seconds, and display it either as words on a screen or as stunning computer graphics.

A fundamental change is taking place. In our day-to-day life, we no long rely on ink on paper to record information. Instead, we count on computer systems which record information digitally on magnetic tapes, on compact disks, or on computer memory chips. When you go through a supermarket check-out line or use an automatic teller machine, you rely on computer records, not paper records.

The volume of data handled by many computer systems would be impossible to print out on paper. The stunning pictures from the Voyager mission to Neptune represent over a trillion bits of data. The Internal Revenue Service has over ten trillion bits of information on American taxpayers. For comparison, the entire Encyclopedia Britannica contains only about ten million bits of information, a million times less.

It is easy to see why we cannot only rely on the printed page anymore. Without advanced computers and computer networks, we could not possibly store and process all that data.

The rapid advances in the field of telecommunications are partly responsible for this trend away from paper. Today we rely less and less on the mail to carry printed information from point A to point B. The information can now be detached from the paper it is printed on. We put a printed page in a fax machine; the information is scanned and sent across the country, leaving the paper behind.

In many offices, paper often never enters the picture. Documents are composed on a computer, distributed electronically, and filed away in computer memory and on magnetic tape. A document may get printed out, but more and more we keep the document on the computer and throw away the paper.

Accelerating this change has been the rapid evolution of computer networks. Today's networks can transmit 50 pages of text from coast to coast in a second. You can't do that with paper. The National Science Foundation runs NSFNET which soon will be able to transmit 45 megabits of data a second. That's fifty typewritten pages in a second. The legislation we will be discussing today would create a National Research and Education Network that would be more than sixty times faster than that. This national network will be able to transmit 3 billion bits of data every second by 1996.

Such speeds are necessary because of another fundamental change in the way we handle information. Just as computers are replacing paper, images are replacing words. More and more, we rely on images to convey information.

In July, the Science Subcommittee held a hearing on visualization -- new technology which uses computer graphics to sort out the billions of bits of data being produced by supercomputers. We heard how scientists are using supercomputer models and computer graphics to study thunderstorms, design better airplanes, and find oil.

We all know that a picture is worth a thousand words, and that is definitely true in computing. In general, images require a thousand times as much data as text. And that's why we need the capacity offered by a 3 gigabit network.

This shift from paper to computers, and from words to pictures, will profoundly change many of our institutions. Our concept of a library will have to change with the technology. Our society has traditionally used libraries to store our information - on paper, and in some cases, on microfilm.

But computer technology offers a new approach, something called the "Digital Library." Books would be stored digitally on magnetic tape, compact disks, or on computers. Using a computer, library patrons could access information from the books in the library's collection. Thousands of books could be on-line, available at the touch of a button. The technology already exists to do this. Today you can buy a single compact disk containing the entire Bible, an unabridged dictionary, the complete works of Shakespeare, and a half a dozen other reference books. Not only is it extremely compact and portable, but you can use a computer to search the disk and find a particular phrase far faster than you could by hand. Later in this hearing, we will hear about commercial products offering hundreds of magazines on a single compact disk.

Of course, a digital library would not really have walls. With computer networks, people around the country could access the digital information on-line. In the future, a child in rural Tennessee could use his or her personal computer to get access to much of the Library of Congress. A single digital library could serve the entire nation.

We are already seeing the start of a system of national digital libraries. The Library of Congress maintains a national computerized card catalog which thousands of libraries nationwide access over computer networks. The Library of Medicine distributes medical information from its databases to doctors around the country using its MEDLINE system. The U.S. Geological Survey is developing the technology needed for a national digital map library. One of our witnesses today will provide a demonstration of this technology.

Last May, I introduced S. 1067, the National High-Performance Computer Technology Act of 1989 to promote the development of such national digital libraries. That legislation calls for the development of a 3-gigabit National Research and Education Network. Connected to that network will be digital libraries of scientific information. Climate data, Landsat images, economic data, and journal articles would all be available over the network.

Scientists and engineers throughout the country are developing many of these databases. As happens so often in computing, the newest, most advanced computer applications are put to use first in the research community. The supercomputer graphics that scientists are using today give us a taste of what TV will be like in ten years. Likewise, the computer networks and scientific digital libraries being built now will lead the way to development of other networks and digital libraries in other fields -- in medicine, in law, in elementary and secondary education.

Of course, developing these digital libraries will not be easy. The switch from paper to computers is a drastic one. This hearing will allow us to explore some of the consequences of such a change and some of the problems we will have to face.

We will need to develop reliable, easy-to-use technology that allows users to browse the mountains of data in digital libraries, and to pull out what they want, using computer networks.

We have to find ways to ensure that information stored electronically is as secure and as durable as the printed page. We have to develop standards so that data stored on one computer system can be combined with data from a different system.

We need to find ways to protect the integrity of data in digital libraries. A data set is worthless if you can't be sure that it has not been altered. Authors won't want their work stored electronically if they cannot be sure it won't be plagiarized or altered without their permission.

If books and journals are going to be stored electronically, there needs to be a way for an author to protect his or her copyright and collect royalties. We will hear later about innovative ways that are being developed to do that.

The research community and the commercial sector are working on these problems and others. S. 1067 will provide additional funding for these efforts by authorizing $1.2 billion dollars over the next five years for advanced computer software and hardware. Much of that money will go for better technology for handling electronic data.

Getting that funding will not be easy, due to the budget deficits we face. But frankly, we cannot afford not to fund this effort. The federal government spent $63 billion dollars last year on research and development. Investing in digital libraries and a national network would pay for itself by increasing the payoff from our Federal R&D investment.

I was delighted last week, when Dr. Allan Bromley, the President's Science Advisor, released the "Federal High Performance Computing Program," a five-year plan developed by the Office of Science and Technology Policy. The plan is very similar to S. 1067, calling for $1.92 billion in new funding over the next five years for advanced computing. Unfortunately, the President and OMB have not yet agreed to fund the program. But the science agencies are pushing hard to get funds for the program included in their fiscal year 1991 budget requests.

In the meantime, we are moving ahead here in the Congress with this legislation. Advanced computing and networks will help keep America competitive. They will help us face the challenge of global change. They will play a key role in improving our educational system. We must move ahead. The potential of this technology is too great not to.

Senator GORE. And let me make a note about our first witness, who was going to be here. Senator Bob Kerrey was going to be our lead-off witness. He is a co-sponsor of S. 1067 and has taken a special interest in this subject. Unfortunately, however, this is appropriations week here in the Senate as we rush to meet the fiscal year deadline, and Senator Kerrey is a member of the Appropriations Committee, where some critical meetings are taking place this morning.

He has, however, submitted testimony for the record, and it will be included in full.

I am delighted now to invite our first panel to come to the witness table. Our first witness is going to be Joe Wyatt, Chancellor of Vanderbilt University in Nashville, and a computer scientist himself. He is someone who is uniquely qualified to discuss the impact of this new technology on our universities, and a good friend, as well.

Dr. James H. Billington is the Librarian of Congress, and he is accompanied by Henrietta Avram, who is one of the leading experts in the country on these subjects. The Library of Congress is already helping libraries put computer networks and databases to better use.

Dr. Robert Kahn is the President of the Corporation for National Research Initiatives, and a personal friend, who, back in the 1960s, helped develop ARPANET, the first national computer network.

I might say that the subcommittee is just delighted to have all four of you here this morning to help us get off to a good start.

What we will do is put all of your full statements in the record and invite you to summarize the principal points for us. Then we will withhold questions until after the full panel has concluded, and then we will move forward.

Joe Wyatt, thank you for coming and thanks for being my partner in helping to think through these ideas for several years now. I certainly do appreciate your presence here.

STATEMENT OF JOE WYATT, CHANCELLOR, VANDERBILT UNIVERSITY

Mr. WYATT. Thank you, Senator Gore. It is a real privilege to be able to speak on a subject that I think is very important. And I thank you for your continuing interest and expertise in this subject.

More than 30 years ago, as a young aerosystems engineer, I began to participate in the design and development of local area networks as well as national digital communication systems, for use in the aerospace industry.

And later, I participated in similar activities, designing systems for university environment and have participated in most, I think, of the national efforts directed at computer networking.

The reason I mention this is that these experiences have demonstrated clearly to me the power and the efficiency of network access to computer systems and collections of data.

In 1983 I was one of 13 scientists who met at the invitation of the National Science Foundation to develop a plan for making supercomputer power available to research scientists across the nation. Dr. Kahn was one of our advisors.

The five supercomputer centers, which have resulted, required a special and significant appropriation by the Congress, with matching funds from business and state government. They were developed competitively under the peer review system. And, of course, you, Senator Gore, were absolutely critical in the passage of that legislation that provided funding for those supercomputer centers.

And I am pleased to say that these centers now play a critical role in the research and education activity in the United States, as well as the transfer of technological developments to American industry.

The supercomputer centers and modest network improvements associated with those centers were clearly a wise and important investment for this country. Now we have come to the next phase: the expanded availability of supercomputer power to research, education and industry in America, and the development of a national interstate network operating initially at gigabit levels, but with the potential for terabit speeds.

The implications of this advancement for American research productivity and educational effectiveness are profound, in my view, and far-reaching. It is my belief that without this development, first, none of America's other major research investments will benefit the country as much or as quickly.

Two, the education system in America will lose its most valuable single hope for revolutionary improvement in quality and equity.

And finally, the information infrastructure which supports both research and education in America will become hopelessly overburdened and ineffective for both purposes.

It is also my belief that without the kind of support proposed in this legislation, these developments cannot occur on a timely basis and may not occur at all.

You have suggested that I focus on some examples from Vanderbilt, and I am pleased to do so with only a few, in the interest of time.

First, let me say a word about libraries, the central focus in the information activity of any university. A library includes the complete collection of information used in both research and education. This environment is changing rapidly, as you have pointed out. But the core technologies for the library remain the book, the shelf and the catalog.

The most consequential attribute of research in teaching libraries is the rate of flow of new books and information entering the collection. And the value of most academic libraries for both research and teaching depends critically on the flow of new materials and the currency of these materials.

The library that is quick to acquire, catalog and shelve its new items is substantially more valuable to its users than a library that is slow to perform these functions.

At Vanderbilt and at most other universities that I know about, recent investments have been dominated by increased budgets for collections, cataloguing and building space; multi-million dollar investments have been made at Vanderbilt alone. But in a recent study by Vanderbilt's librarian and presented in an international conference in Paris, it is forecasted that the world will produce between 900,000 and 1.1 million new titles per year by the year 2000, if not before.

No single library, certainly not Vanderbilt's, can afford to collect and make available to its users even a majority of these titles, much less all of them.

For more than two decades, national library networks, as you mentioned earlier, have provided a means for searching electronically the catalogues of multiple university libraries. Most of the major research libraries in this country and many around the world are connected to these networks. They are at the Ohio College Library Consortium, the Research Libraries Group, Medline, and other specialized collections. But even all of these facilities combined cannot keep pace with the publication rate in print approaching 1 million titles per year.

Therefore, the future library must be invested heavily in the prospect for electronic information services. With electronic services, the library is no longer limited solely by the book. Electronic services also provide a vehicle for coping with the growth in print by providing on-line access to collections worldwide of bibliographies.

And finally, in some disciplines, scholars, even now, develop originals using word processing equipment implemented through computers and communicate these documents worldwide through electronic means.

In the new world of electronic publications, I believe that individual scholars will develop private libraries of their own in electronic form for their research and teaching. Computer-based tools for managing these personal data collections already allow scholars to gather and manage larger files of documents than they can possibly hold as personal print libraries, and we see the software developing rapidly, along with the hardware, for facilitating this.

In addition, scholars can and will regularly access directly, over national and international networks, the catalogs and files of hundreds of other libraries. Clearly, there can be no robust library of the future for any university without the availability of a national network of the sort proposed in this legislation.

The second example relates to research. From the beginning, the most important scientific applications of digital computers have been simulations in the form of mathematical models of complex physical phenomena that are difficult, if not impossible, to create otherwise.

Larger and faster computers have allowed us to do more detailed simulations over longer periods of simulated time. These advances in computer capacity and speed have enabled the creation of more useful experiments in virtually every field of research and science and engineering. You have mentioned some earlier; molecular studies of matter for the synthesis of new materials and pharmaceuticals, meteorological models for weather systems for understanding storms and longer-range climatic changes, geological models to explore subsurface formations for oil and gas—the list is really endless, because it continues to grow.

In the report which proposed to the Congress the funding of the five national supercomputer centers in 1983, we reached several conclusions, two that I think are especially pertinent to today's discussion.

First, we concluded that America's future preeminence in basic research would depend on the development and use of high performance computer systems.

Second, we also concluded that the effective use of such systems would depend in large part on the development of a very high-speed national computer network which would provide research scientists convenient and continual access to the high performance of supercomputers on the network and to collections of data available there.

As the national supercomputer plan developed, Vanderbilt was presented with the decision of whether to develop a proposal to be one of the five supercomputer centers, or to develop a strategy in which we would develop other computer resources on campus and become a remote user of the supercomputer centers by means of network access. That was true of many, many universities, and for a variety of reasons, particularly the capital investment requirements, we chose the latter course.

In this strategy, we are developing a high-speed local area network connected to all Vanderbilt buildings, including dormitories. In addition, specialized computer resources proposed and developed on a case-by-case basis at Vanderbilt continue to be installed in the various schools, departments and laboratories of the university.

Finally, we are developing specialized research laboratories which connect to the network in such a way that campus users and users from around the country can both access information produced in these labs and can contribute experimental data themselves.

One such example is the National Free Electron Laser Laboratory now being completed at Vanderbilt. We are one of the very few American universities which will house a free electron laser, a piece of equipment with potential use in research ranging from laser surgery to physics of surfaces, all leading to a better understanding of the effect on matter of the beam.

When it becomes operable later this year, the free electron laser will generate a large collection of data which will accumulate from the experiments. The evaluation of most of these data will require the use of computer systems—in some cases, supercomputer systems—and both the raw data and summarized results will be stored in computer data bases.

These results will be available to other researchers around the Nation by means of connection to the local area network, which in turn is connected to national networks, and the effective use of this data will depend in large part on the capacity of the national network being in the gigabit range.

Researchers who come to the facility and conduct experiments can then go back to home base and continue their work with the free electron laser experimental data by means of the network.

There are literally hundreds of other examples that I could give, and I will not do so in the interests of time, but virtually every one of these research activities will benefit directly from higher network speeds and direct access to high performance computer systems.

Now to education. There are immediate and important advantages to a national network for improving the instruction and educational opportunities for teachers and students, regardless of their level of study.

At Vanderbilt, at the college level, this month we opened one of the most advanced multimedia classroom facilities in the Nation. Each student desk has an advanced computer work station, the computers are connected together to access information at a high rate of speed from either data stored in a large disk storage available in the classroom or across the campus network.

The instructor can control all of the computers and send and retrieve information from an individual student's computer or to all students' computers. In addition, the instructor controls a six foot by eight foot projection screen, various access to video cassette recorders, disk players, cameras and other devices that interface with the local area network and national network.

A host of classes—more than a dozen—from calculus to psychology are now being taught in this classroom this semester.

Our engineering school is taking this electronic classroom one step further, to a distributed classroom. In this model, the master classroom, similar to the one I just described, would be connected to remote classrooms over a broad geographic area so that especially effective teachers can reach more students, regardless of their location.

The system allows the instructor to retrieve what is called a window of information from any individual work station anywhere, and to send information there, including a television image of the instructor in a small window on the screen.

A prototype distributed classroom, the first one, will connect classrooms at Vanderbilt with classrooms at NASA's Marshall Space Flight Center in Huntsville, Alabama, and will be capable of courses at the graduate and the undergraduate level.

We know that these are important improvements for Vanderbilt education, but we also know that the most far-reaching prospect for this classroom is to extend the distributed concept into the high schools, where particular course deficiencies—take calculus, for instance—can be treated early.

Only a network like the one proposed here will suffice for this purpose, and the stakes grow higher as we contemplate a shortage of mathematics-trained high school teachers which promises to be even worse in the future than it is today.

I think the network can also contribute to the solution of problems in education long before high school. Just to give you one example, Vanderbilt researchers have developed adventure stories on interactive video disks that are designed to change the way K through 12 students approach mathematical problem-solving.

Instead of traditional word problems, students are challenged to first formulate problems and then solve them with clues spread throughout the adventure. In the first episode, students must determine whether a fictional character, Jasper, can get his boat home before sunset without running out of fuel. The second episode, using an ultra light aircraft to rescue a wounded bald eagle, the problem is to figure out the quickest way to get the wounded eagle to the vet.

The interactive nature of this technology allows students to search for information they may have missed, and the accompanying database allows teachers to deal with ancillary subjects, such as the history of river travel, or the geographic location of major rivers in the world. But the idea is that the students have to formulate the problem from clues in the material, and then gather parametric data from the material to solve the problem.

The first video disk has generated an enormous amount of excitement in about 100 classroom settings of fifth and sixth graders. With a supercomputer network, this instruction could be available to even the most disadvantaged school district. Moreover, teachers would be able to share, via the network, the technology and new data so that other teachers benefit, and future adventures can include suggestions from that hands-on use by all teachers.

All of the examples demonstrate that the National High Performance Computer Technology Act of 1989 is a keystone for the future structure of America's educational system and its multifaceted research enterprise, and I believe it is crucial to the development of more effective linkages between American business and education. We cannot remain competitive as a Nation without improving these linkages, and soon.

The President's science adviser's new report and recommendations are consistent, in my view, in every important respect, with Senator Gore's National High Performance Computer Technology Act of 1989.

Without the kind of Federal Government support proposed in this legislation, the critical developments necessary for this country to maintain world leadership cannot occur in time. I urge passage of S. 1067.

Thank you.

Senator GORE. That was a terrific statement. You will not be surprised to hear my assessment of it. I appreciate that very much.

I was listening to you talk about making especially effective teachers available to other classrooms through this network—for teaching calculus was your example. Anybody who has seen the movie, "Stand and Deliver," can just savor the prospect of a Jaime Escalante being able to inspire and energize many, many thousands of students instead of the few in his classroom.

In any event, I appreciate the statement.

Dr. Jim Billington is accompanied, as I said earlier, by Miss Henriette-Avram, Assistant Librarian for Processing Services. Dr. Billington, thank you very much. Please proceed.

STATEMENT OF HON. JAMES H. BILLINGTON, LIBRARIAN, LIBRARY OF CONGRESS; ACCOMPANIED BY HENRIETTE AVRAM, ASSISTANT LIBRARIAN FOR PROCESSING SERVICES

Dr. BILLINGTON. Thank you very much, Mr. Chairman and members of the Subcommittee. I appreciate the opportunity to be here today with my distinguished colleague, Henriette Avram, the Assistant Librarian for Processing Services, to comment on S. 1067, a bill to provide for a coordinated Federal research program to ensure continued U.S. leadership in high performance computing.

You have already heard a very eloquent statement of the broad picture. My comments will be limited to the role of the Library of Congress and the role other research libraries should play, and can play, I believe, in such a national computer highway.

I will describe some of the freight, so to speak, that can be carried on this highway, as well as some of the ways the transportation can move.

As you know, the Library of Congress has for over 20 years made available in tape mode machine-readable bibliographic data describing materials received in the Library of Congress.

We estimate that this data saves libraries around the country at least $360 million a year annually in cataloging costs—an enormous and not always fully-understood direct service to the Nation's libraries and schools by the legislative branch of this government.

In the last ten years, we have also been providing on-line service for cooperative cataloging projects that benefit the Library of Congress and other libraries, so there is the beginning of a network in this area as well.

Our most recent project, announced just last week, is a pilot program to put on-line to the Library of Congress data bases 14 libraries of different kinds and in different communities throughout America. For the first time these libraries will get as well as access to status of bills in Congress, copyright and research referral services, direct access on line to the Library of Congress' automated data on bibliographic records.

With 88 million items in the library—we receive 31,000 items a day, Mr. Chairman, from all over the world in all languages and in all formats—we represent by a considerable margin the largest collection of recorded information and knowledge ever assembled in one place in history. Most of it is right here on Capitol Hill in the three buildings adjacent to the U.S. Capitol.

The library represents the Nation's most important single resource for the information age, and the proposed establishment of a National Research and Education Network would give an immense boost to making this material accessible to the entire citizenry.

Now, it is obvious that the network envisioned in S. 1067 could be very important in making available Library of Congress resources to more scholars and researchers around the United States, as you indicated in your own introductory remarks.

The library has important databases that it develops as search guides to information collected from all over the world on every subject except clinical medicine and technical agriculture. These fields have their own national libraries, as you know.

The library's collection of foreign language materials—more than two-thirds of its materials—are unique to the country and, in many respects, to the world. We have the largest collections of Slavic, Hispanic, Germanic, Chinese, Japanese, Korean, and Arabic materials, outside of the countries of origin, in the world.

This information, combined with the experts we have who catalog and process it, who in effect develop a stored knowledge of the bibliographic production of the entire world, is a major international resource for scientists, economists, policymakers and other scholars in this country.

Yet, this is an almost entirely unused resource for the productive economy of this nation. We have an enormous array of scientific and technical reports, as well as published materials in all languages and formats, enriched by our recent acquisition from the University of Chicago of the National Translation Center, with bibliographic information, or full text translations, on a million technical translations.

Senator GORE. If I could interject a point here just to illustrate that, how many fluent Russian speakers do you have in the Library of Congress, compared to the State Department?

Dr. BILLINGTON. We have 230 Russian-speaking people on the payroll of the Library of Congress.

I am not sure what the State Department figure is, but I would be very surprised if you could equal our foreign-language speakers in terms of fluency, because ours are people who are reading and describing the intellectual production of the world in the cataloging under Ms. Avram's brilliant leadership.

We have between 120 and 130 Japanese- and Chinese-speaking persons. These are bilingual people who can read and converse, so it is quite a resource.

Senator GORE. I apologize for interrupting. I wanted to understand that point, because you were talking about what a tremendous resource this is for this country; that we do not make adequate use of in this country.

Excuse me for interrupting. Go ahead.

Dr. BILLINGTON. It is a resource presently channeled very heavily into generating this fundamental bibliographical data on which the thousands of libraries of the country depend, and on which some of the vendors also largely depend. But it is a resource of even greater ultimate potential, in terms of the kind of broadening horizons that we are talking about this morning.

These automated bibliographic files assist users in identifying the existence and location of all this information. We are working also in the image data area with optical disk and video disk projects.

In addition to books and periodicals, the Library's collections encompass, of course, maps, charts, motion pictures, prints and photographs, and computer software, to name only a few of the many formats.

We have specialized reading rooms for all of those formats and several others. Opening up a broader network gives the possibility, by capturing collections in electronic form, of a "library without walls," as you have mentioned, for those materials, which are not available elsewhere. A good percentage of our collections, in fact, are not available elsewhere.

Now we believe that we can make such information and knowledge increasingly accessible to users beyond the Library of Congress premises and that a sound, high-capacity networking system is very much needed to accomplish this. Image data, as you know, is very high-volume data. The network proposed in S. 1067 would appear to be a natural vehicle for disseminating our electronically recorded resources.

Our video disk project, in recent years, has successfully captured images of a portion of our graphics collections in machine-readable form, which could become a significant resource during the next several years for researchers across the country and also for educators, along some of the lines that my distinguished predecessor in his testimony has indicated.

American Memory is a major new Library of Congress initiative, using optical disk and other advanced technologies to disseminate unique elements of the library's vast and variegated collections of American history, culture, folklore, and the like.

The proposed high-speed, high-capacity National Research and Education Network could greatly facilitate the wider availability of all our resources to a wide variety of users in the education sector, in the economic sector, as well as in the library and research community.

During the last decade, Mrs. Avram has represented libraries on the EDUCOM board of trustees to ensure that any proposed national network will meet the needs of libraries as well as that of the research community.

The proposed network must incorporate a consistent family of standards and techniques. This is a lesson learned by the Library of Congress as it began the kind of standardization of cataloging necessary for the entire development of our unmatched national library system.

In cooperation with other libraries, the Library of Congress has devised national standards for cataloging, and this single network should fit the needs both of those creating and exchanging information, and of those searching for the information for research purposes.

In other words, both educational institutions and libraries need to be operating a system of unified national standards.

And this brings me to Title III of your bill, the National Information Infrastructure. The National Science Foundation is directed to coordinate, with federal agencies, the development of a national science and technology information infrastructure.

We believe that the Library of Congress, as the national library and as the largest information resource in the federal establishment, should play a prominent role in the development of such an infrastructure—particularly in the setting of standards where we have long held the natural leadership role.

Now, Mr. Chairman, in your invitation to me to testify, you asked that I also address the issue of copyright. My colleagues in the Library of Congress' Copyright Office believe that in most instances, an author whose works are available on-line, over a network, can protect himself or herself through contractual agreements with vendors and libraries and through copyright license agreements.

It is a complicated area. I discuss it a little more in my extended remarks, but let me just say now that, currently, database vendors enter into contract with purchasers of their services that include restriction of access and use. Part of the fees paid by purchasers typically reflect copyright royalty payments for use of copyright materials. But abuses can occur, and this issue will need much further study.

The question of fees for information on a national computer network directly relates to another subject. This other subject is discussed at length by educators and librarians, but on the whole, is addressed by others only rather peripherally—the area that is broadly referred to as national information policy.

Now, as you know, congressional committees with jurisdiction over federal information and its distribution have held hearings in recent years. But no national policy has yet been fully articulated.

While S. 1067 is a step forward, additional efforts will probably have to be made in this area concurrently. We must, for instance, recognize the rights of our citizens to obtain this Federal information, but we must also address carefully and equitably—and then face up to directly—the issue of user fees for information gathered both at the national and the local levels.

Some compromises will be necessary to accommodate these competing interests, those of the users and of the creators, if we are to avoid becoming a nation of information haves and have nots. Information must not be too expensive and it must be widely shared if the talents of this country are to keep us creatively alive and internationally competitive.

The Library of Congress and the library community of this country, which is an enormous and variegated resource, are, I believe, united in wanting to help develop a national research and education network. And our citizens have a right to expect that irreplaceable intellectual resources in our nation's libraries be made ever more widely available through a national information network.

After all, the creation of the whole network of libraries in this country was an expression of the importance attached in the industrial age to developing this opportunity and this access to information for all our citizens at a time when they did not have as much access through other institutions as they might. And this is very much a part of the same impulse.

Ben Jones, for instance, says that the industrial revolution was the period in which the powers of the body were extended through machines and things that enabled the body to do more. We are now living in the mind-extending age. The information age is one where the horizons are infinite. And it is the only area, Mr. Chairman, in which the horizons of freedom are truly infinite, because physical constraints necessarily intrude in other areas.

So, above and beyond the very practical needs which you are addressing here, you are also pointing to our cherished ideal of freedom and of opportunity, really, for all our people, and in a way which is ultimately not even competitive with the peoples of other countries.

So, we want to thank you for this opportunity, Mr. Chairman, and to compliment you on your vision in addressing what I, and I think, most of my colleagues in the library community consider an important national priority.

We at the Library of Congress are anxious to assist you in this effort and Ms. Avram is here with me to answer any of your committee's more detailed questions that you may want to put before us.

[The statement and questions and answers follow:]

Statement of James H. Billington
The Librarian of Congress
Subcommittee on Science, Technology, and Space
Committee on Commerce, Science and Transportation
United States Senate
September 15, 1989

Mr. Chairman and Members of the Subcommittee:

I appreciate the opportunity to be here today with my colleague, Henriette Avram, Assistant Librarian for Processing Services, to comment on S. 1067, a bill to provide for a coordinated Federal research program to ensure continued U.S. leadership in high performance computing. My comments will be limited to the role the Library of Congress and other research libraries should play in such a national computer highway.

As you know, the Library of Congress has for over twenty years made machine-readable bibliographic data on materials received in the Library of Congress available in tape mode. This data, that is of enormous benefit to the nation, saves libraries around the country some $360 million in cataloging costs annually. In the last 10 years we also have been providing online service for cooperative cataloging projects benefiting the Library of Congress and other libraries.

Our most recent project -- announced just last week -- is a pilot program to put online 14 libraries of different kinds throughout America -- state, public, government and university -- offering direct access to the Library of Congress' automated data on bibliographic records, status of bills in Congress, copyright, and research referral services.

With 88 million items in the Library we have the largest collection of recorded information and knowledge ever assembled in one place here on Capitol Hill. The Library of Congress represents the nation's most important single resource for the information age. The proposed establishment of a National Research and Education Network would give an immense boost to the access of this material and allow the Library of Congress to provide to the country much more of its unequalled data and resources which can now be obtained only by visiting Washington.

Already, in our modest nation-wide pilot program for online access to our data, the first question of many participants is "Can we use the existing National Science Foundation network for access to Library of Congress data?" This indicates the interest in a major national network.

It is obvious that the network envisioned in S. 1067 could be very important in making available Library of Congress resources to more scholars and researchers around the United States. The Library of Congress has important data bases that it develops as search guides to information collected from all over the world on every subject except clinical medicine and technical agriculture. The Library's collections of foreign-language materials -- two-thirds of its collections -- are unique to this country. We have the largest collections of Slavic, Hispanic, Chinese, Japanese, and Korean materials outside of the country of origin in the world. This information is the major international resource for scientists, economists, policymakers, and other scholars in this country.

Moreover, the Library of Congress has recently acquired from the University of Chicago the National Translation Center which is an international depository and referral center for helping users locate unpublished translations of foreign-language literature in the natural, physical, medical, and social sciences. The Center's files contain information on the whereabouts of 1 million translations, of which 400,000 full text translations are held directly by the Center.

The Library's automated bibliographic files assist users in identifying the existence and location of this information which is essential to those wanting access to these rich resources.

We are working in the image data area with optical disk and video disk projects. In addition to books and periodicals, the Library's collections encompass all forms of materials -- maps, charts, motion pictures, prints and photographs, and computer software, to name only a few. Once these materials are captured in electronic form, we envision a "library without walls" for those materials which are unavailable elsewhere. We believe that we can make such data increasingly accessible to users beyond the Library of Congress premises. To do so a sound, high-capacity networking system is needed. Image data, as you know, is very high-volume data. The network proposed in S. 1067 would appear to be a natural vehicle for disseminating our electronically recorded resources.

Increasingly, data will be coming to the Library of Congress in machine-readable form or be generated at the Library in this form. Major educational institutions are also creating electronic information.

A significant asset at the Library of Congress is our graphics collections -- prints, photographs, drawings, maps, charts, and essential scientific graphs. Often the only copies in existence are at the Library of Congress. Our video disk project has successfully captured images of a portion of this material in machine-readable form, which could become a significant resource during the next several years for researchers across the country. "American Memory" is a major new Library of Congress initiative that will use optical disk and other advanced technologies to disseminate unique and variegated elements of the Library's vast collections of American history and culture throughout this country and the world. It will put Library of Congress collections in libraries and schools around the U.S.

The proposed high speed, high capacity National Research and Education Network could greatly facilitate this broadening access to all our resources.

During the last decade Mrs. Avram has represented libraries on the EDUCOM Board of Trustees to insure that any proposed national network will meet the needs of libraries as well as that of the research community. The proposed network must incorporate a consistent family of standards and techniques -- just as the Library of Congress, in cooperation with other libraries, has devised national standards for cataloging. This single network should fulfill the requirements both of those creating and exchanging information and of those searching for the information for research purposes.

This brings me to Title III of your bill, the National Information Infrastructure: the National Science Foundation is directed to coordinate with Federal agencies the development of a national science and technology information infrastructure. We believe that the Library of Congress -- as the national library and as the largest information resource in the Federal establishment -- should play a prominent role in the development of such an infrastructure.

In your invitation to me to testify today, you asked that I address the issue of copyright. My colleagues in the Library's Copyright Office believe that in most instances an author whose works are available online over a network can protect himself or herself through contractual agreements with vendors and libraries and through copyright license agreements. Computers have a record-keeping and transaction tracking capability that could be used to establish and account for user fees for specific services and works. These fees could include a royalty payment for use of copyrighted material.

Currently, data base vendors enter into contract with purchasers of their services that include restrictions of access and use. Part of the fees paid by purchasers typically reflect copyright royalty payments for use of copyright materials. But abuses can occur. This issue will need much further study.

The question of fees for information on a national computer network directly relates to a subject that is discussed at length by educators and librarians but which is addressed by others only peripherally -- that is, "national information policy." As you know, congressional committees with jurisdiction over federal information and its distribution have held hearings in recent years, but no national policy has been articulated. While S. 1067 is a step forward, additional efforts must be made to develop a policy on Federal information if the United States is to remain competitive.

We must recognize the rights of our citizens to obtain this information. We must also face the issue of user fees for information gathered both at the national and local levels. Some compromises will be necessary to accommodate these competing interests -- the users and the creators -- if we are to avoid becoming a nation of information haves and have-nots. Information must not be too expensive, and it must be widely shared if the full creative talents of the country are to keep us creatively alive and internationally competitive. Yet to some extent user fees must support the creation and dissemination of this information.

The Library of Congress and the national library community want to help develop a national research and education network. The major information resources for this nation reside in the Library of Congress and other research libraries across the country. These collections have been assembled in many cases with Federal, state, and local revenues. Our citizens have a right to expect that these irreplaceable intellectual resources be made ever more widely available through a national information network, such as that envisioned in S. 1067.

In conclusion, Mr. Chairman, I want to compliment you on your vision in addressing what I consider an important national priority. The Library of Congress is anxious to assist you in this effort and Mrs. Avram is here with me to answer any of your Committee's questions.

QUESTIONS OF SENATOR GORE AND THE ANSWERS THERETO

LIBRARIES AND THE NATIONAL RESEARCH AND EDUCATION NETWORK

1. Why did the research community and the library community go separate ways?

Libraries have a tradition of sharing, and they depend on standards in order to share information. Furthermore, various networking experiments in the 1970s, demonstrated the need for a communication architecture which would be system and application independent.

Thus during the development of the national library network in the late 1970s and early 1980's the library community adopted communication protocols based on emerging international standards (i.e. the protocols of Open Systems Interconnection specified by ISO, the International Standards Organization). The guiding philosophy was that adherence to emerging standards would reduce the long-term cost to accommodate new systems and new applications, as well as future standards to be developed in accordance with this emerging communication architecture.

On the other hand, these international standard protocols have matured slowly. In the academic and research community in the U.S., impatience with the lack of an operational networking solution has led to the use of an interim communication architecture, based on communication protocols used by the defense department (the DARPA suite of protocols).

2. Describe efforts underway to link the networks.

The library community and the academic community each has valuable resources to offer the other. Libraries have databases of interest to researchers. Academic networks have communication bandwidth to share -- and telecommunication costs are always of concern to libraries. Consequently, there has been extensive collaboration between these two communities to keep library networking and academic networking from branching into incompatible and irreconcilable directions.

Pilot projects are under consideration to link academic workstations and libraries over NSFNet. NSF has stated that NSFNet will migrate from the DARPA protocol suite to OSI, the international standard communication architecture, perhaps within the next few years. These pilot projects would employ interim migration steps, intended to facilitate migration, while at the same time allow the systems to avail themselves of the rich functionality and applications provided by the standard protocols.

Some of the pilot applications which have been discussed include:

- The University of Maryland contributing bibliographic information to the Library of Congress;
- a scholar at a workstation searching a library database, such as OCLC, RLG, or the Library of Congress;

- a Library of Congress researcher searching a remote database.

NYSERNet, the New York State Education and Research Network, one of the regional networks on NSFNet, is planning to provide a gateway between the two communication architectures which will allow an academic workstation to search an information provider such as OCLC. NYSERNet is also implementing one of the key protocols of the international standard suite, namely the directory protocols, which will dramatically facilitate access to information by making it easier for a user to locate desired resources.

3. Should S. 1067 explicitly require that libraries be connected to the proposed network?

No. It should not be required that libraries be directly connected to the proposed network, but it must be explicitly required that the network interconnect with other existing commercial networks.

Libraries select and purchase services according to their needs, and must be allowed to continue to do so. Thus it is important that other networks remain viable and motivated by competition to provide the best possible service. Even if libraries become heavy users of NSFNet, it is unlikely that all libraries will use it. However, it is critically important that universal connectivity be achievable, so that libraries on different networks may communicate. It is important that the end result be one logical network.

In the past, commercial networks were hesitant to implement internetworking capability, although the technology clearly existed. This was a major obstacles to interlibrary communication. The proposed network will likely provide access to wide variety of information resource, and this will provide the economic incentive for competing networks to interconnect with the proposed network. Legislation should explicitly encourage and support these interconnections.

4. Do libraries really need a gigabit network, or would a slower network do the job?

Libraries do not need a gigabit network until libraries are "digitized". When libraries begin providing full-text on-line, then a significantly larger capacity network will be required. Until then, a slower network can be adequate.

However, "gigabit" refers to raw capacity, which is only one of several performance, integrity, and functional characteristics of a network. For example, for one application, transit delay or connection set-up might be more important than capacity; for another application, resilience might be important.

In other words, a network will not automatically fulfill library requirements simply because it is a gigabit network.

5. If libraries provide books and journals on-line over NSFNet, would it swamp the network?

It is unlikely that library data, including full-text, would overload the network based on volume alone.

It is expected that such a network would need to grow as volume increased, and the technology will be advancing accordingly.

6. If libraries become heavy users of NSFNet, how much will the added capacity cost, and who will pay? NSF? The Library of Congress? Someone else?

If libraries become heavy users of NSFNet, the cost would depend on what applications are being run.

However, the library community would not bear the cost of the added capacity. Libraries would most likely play the role of information provider much more than user. That is, far more data will be generated as a result of libraries providing service to the community at large, than data generated as a result of services provided to libraries.

It is envisioned that the federal government would provide initial start-up funding plus seed money for pilot projects. After this initial period, the network must be self-sustaining, and compete with -- and most importantly -- interconnect with, other long-haul networks.

As a self-sustaining network, the users pay through fees. NSF and LC will not pay.

DRASTIC CHANGES IN STORE

7. What will happen to the roles of the traditional institutions of scientific communication -- libraries, publishers, professional societies and so on -- when the infrastructure proposed in S.1067 becomes fully implemented, when we have gigabit networks, electronic publication, digital libraries, and so on?

The "traditional institutions" will change along with the material, as they have adapted to microform, external database access, automation, etc. The initial network implemented as a result of S.1067 will primarily affect scientific material. Libraries and publishers will continue to produce and handle more print than electronic material for many years, but the proportions may slowly change from print to electronic.

Publishers will continue in their role as agencies that make the electronic data available, both online and in other formats. They will still be needed by authors to handle the sales and distribution function.

Libraries will still function to facilitate access to data. The library role today is not just to store data but to help users identify the information that is relevant to their needs and access it. That assistance will be even more important with the advent of electronic access, when information users may be flooded with additional unfiltered data, and will pay for each access. The Library will continue to provide researchers and students with a focal point for accessing multiple electronic publications. Some of these publications will reside in local computers, others will be remotely accessed. As more data is published in electronic form, publishers and libraries will probably work more closely and in some instances the functions may merge.

8. What do you think will be the largest impact on the man in the street?

The man on the street will not be directly affected for many years, but if more timely information exchange improves the U.S. economy, he will reap enormous benefit. Many citizens who are not in the research community are beginning to be effected by the increase in availability of electronic information in their libraries and offices. This project will spur the development of electronic access.

ROLE OF THE COMMERCIAL SECTOR

9. What incentives would encourage the development of privately offered network services?

Access to the Research network will likely be the only incentive required for development of privately offered network services. The broad reach of the Research network will create a large potential market for services that the U.S. information industry, among others, is likely to be eager to gain access to. Some of the major commercial bulletin board and data base companies have long shown an interest is connecting to academic networks, and computer-related vendors are interested in using the networks to enhance their customer support. Research and educational institutions themselves may be tempted to offer innovative support services on a fee-for-service basis.

10. What do you think of the idea of allowing commercial database companies to use federal networks like NSFNET or the proposed 3-gigabit network? Should the Federal government charge them? If so, how much?

The Research network should allow researchers, students, and universities to be able to access commercial databases as well as exchange messages with each other. Electronic bibliographic finding aids, full text journals, reference information, and books may primarily be available from commercial sources. The usefulness of the network would be severely limited if these sources could not be accessed.

It should be noted that the commercial services would still serve other users, such as businesses, over private-sector networks. The research network would not displace the bulk of the private sector business.

If the Research Network is used by scientists to access bibliographic or full text databases, the scientist should pay data base charges to the commercial service and telecommunications charges to the Research Network.

If businesses access Federal databases over the Research network, they should pay access charges to the Federal agency and telecommunications charges to Network.

The telecommunications charges should be a function of costs as the network should be self supporting for daily operations.

COST OF THE NATIONAL DIGITAL LIBRARIES

11. What kind of existing federal data bases are available on NSFNET?

Scientific databases such as containing satellite data and medical data bases produced by the National Library of Medicine.

12. Which federal databases would you most like to see on the network?

Every meaningful and externally useful (and unclassified) federal database of statistical, textual and bibliographic information should be made available if possible. This does not include administrative data bases such as personnel and pay, customer files, library charge files, classified information, etc. There are databases relating to, for example, census and calculation of GNP that might be useful. The Library of Congress and National Agricultural Library bibliographic files would be basic.

Government researchers who need to exchange classified data should be able to use the network if adequate security can be provided.

13. Are there any technological hurdles to making large federal databases accessible over the networks?

The major barriers are associated with costs and standards, rather than with particular technical problems. Should any databases on the networks contain classified or private data, there would need to be additional security measures employed as well. There are a number of solutions to this problem, such as separation of systems, but they are often awkward and inefficient to implement.

In the case of data received by the Library in machine-readable form, it may be feasible to enter into contracts with the database vendors to obtain the necessary copyright permissions to permit further dissemination to remote points.

14. What are the primary hurdles to do this? Are they technological or legal?

The primary hurdles are the cost and time required to input existing analog material into digital storage-retrieval media, and the time required to obtain copyright permissions to permit the transformation of the vast holdings of the Library of Congress in to machine-readable formats for dissemination outside the Library.

15. Is it practical to think about digitizing the Library of Congress? How much would it cost? How many people would be required to do the job?

It would cost at least $893,000,000 (1989 dollars) to digitize the Library of Congress. This calculation is based on the following.

Textual collection of 15,000,000 books would cost $.19 per page to digitize (which includes workstations with 5-year life expectancy and labor). The time to complete would be 25 years working 24 hour days. The visual image collection of 11,000,000 photographs, prints, and drawings images would cost $3.29 per image.

This figure does not include maps, which present special capture problems. It also excludes items such as phonodiscs, glass plates, slides, motion pictures, and similar media.

It is not really practical to digitize the whole retrospective collection indiscriminately. Rather a program of capturing current material, older material easily identified for scientific importance, and selected brittle material should be pursued.

16. Publishers are understandably worried about the idea of putting the Library of Congress on-line. Are they universally opposed to the idea? Or do some see electronic publishing as an opportunity?

Publishers are worried about the use of their copyrighted works in computer networks. Some publishers do see electronic publishing as an opportunity to license use of their works in new markets. In implementing the optical disk pilot project, the Library of Congress worked closely with publisher representatives and obtained permission to use the works as part of the project. In return for royalty-free license, the Library agreed to share data about the project with publishers. The data can be used by publishers as market research regarding the potential market for computer retrieval of their works in libraries. Publishers will continue to see electronic publishing as a threat as well as an opportunity unless they are reassured that the libraries that use their works in computer networks will enter into appropriate copyright licensing agreements.

COMPETITIVE AMERICA

17. What do you think are the main benefits for the economy that digital libraries will provide?

Digital libraries have the potential to expand access to information to a much broader audience, improve the timeliness of the information transfer process, enable more effective searching for information, and reduce duplication of effort for both libraries and individual researchers. These capabilities contribute favorably to the economy by improving research productivity. While it is difficult to give precise data for the value of improved access to information, anecdotal evidence provides some indicators. For example, it may take only 10 to 15 minutes to search electronically 20 years of Chemical Abstracts compared to the hours it might take to do so manually. In addition, any new value-added services that might develop as a result of having digital libraries would contribute to the economy.

18. Of all the things that are supposed to help us stay competitive, where do digital libraries rank?

Digital libraries are one element among many that can help the United States remain competitive. Therefore, it is difficult to isolate its specific importance compared to the numerous other factors involved. The attributes described above indicate how digital libraries can improve U.S. research and development productivity and thereby improve the U.S. competitive position.

19. What barriers do you see to creating digital libraries? How can we overcome them?

The most obvious barrier is lack of resources for personnel, equipment, and access machines. This lack can be initial and also ongoing, for as technology changes, machines and systems will need to be replaced. The chief ways to overcome this will be to increase gradually the library facilities as the amount of digital material increases and to place a strong emphasis on the use of national and international standards in all aspects of the process. Standards are a key factor for stabilizing the environment.

20. What is the danger of not building digital libraries and a national high-speed network?

Our scholars may be left behind first by their Japanese and then by their European colleagues regarding timeliness of exchange and use of information -- information produced by other scientists in the U.S. and ideas coming in from abroad.

Senator GORE. Thank you very much, Dr. Billington, and we will have questions when our final panelist has completed his statement.

Dr. Kahn, I introduced you earlier, but you are President of the Corporation for National Research Initiatives and one of the pioneers in establishing computer networks. And, as I said to Joe Wyatt, I want to thank you for helping me think through some of these things and helping the subcommittee over the last few years. Please proceed with your statement.

STATEMENT OF ROBERT E. KAHN, PRESIDENT, CORPORATION FOR NATIONAL RESEARCH INITIATIVES

Dr. KAHN. Thank you very much, Mr. Chairman and members of the subcommittee.

First of all, I do appreciate the opportunity to appear before you today. It is always a pleasure to be here.

When we established the Corporation for the National Research Initiatives back in 1986, the goal was to foster as well as conduct R&D on a "national information infrastructure." That was a term which was not widely understood by many people at the time, and it really is a pleasure to see it on the Senate's agenda now, because it means that there is a wider appreciation of just what that could mean and where it could lead.

Today, I would like to address my remarks to (1) the National Research and Education Network; (2) its role as part of this national information infrastructure; (3) the digital library system as a concrete example of a piece of infrastructure; and (4) the role that the Federal Government might play in fostering its development.

First of all, I hardly need to say that I am very supportive of your efforts to create a national research and education network; it is clearly important for the nation. And I think you have heard from just about everybody who has testified that this network is a national priority; it is not something that is only of benefit to one group; it is something that can benefit virtually every sector of society. It is a good idea whose time has come, and I certainly hope it passes.

While considerable attention has been paid to the subject of research, not only in your bill, but throughout the departments and agencies of the Federal Government, there has been relatively little attention paid to the educational side of the network. I think this is an area, which is a primary one for the states to get involved in. I plan to return to that issue a little later in my testimony.

I also believe that efforts must go beyond the development of the national research and education network as simply a backbone for communications. I have previously talked about the backbone being what I call the lowest level of the infrastructure. In order to make this technology really available to everybody, however, we must create the higher levels of this infrastructure, of which the digital library system is merely one example.

I have had several interactions with your staff over the past few years about the subject of infrastructure. We have talked about not only the digital library system, we have talked about a national knowledge bank for science and technology, for medicine, or whatever. That might form the basis for some of the educational initiatives. We have talked about a medical exchange, an electronic transaction framework, rapid prototyping; the list is quite extensive. You may recall that from some of my previous testimony, where the subject first came up.

I would like to expand on some of these infrastructure ideas, and particularly, to talk about the process by which we might proceed in this area, and also what the federal role is.

Infrastructure is not a single entity. It is not a thing all by itself; rather, it is more like a label for many separate developments or activities, just as education is a label for many, many different activities, and government itself is a label for many activities.

In order for a development to be a piece of infrastructure, it must be widespread, ubiquitous; it must be shareable in some sense; and it must be relatively easy to use.

I believe the decision to focus on the digital library system is an excellent one. It is one that is very easily conceived by people; they have a natural affinity to it; you can almost see what the immediate benefits would be; and I think it is both important and timely.

But even the digital library system is not a monolithic concept. It has a significant substructure to it. I believe that if we do create a national digital library system, it will be important to involve all the affected parties in developing this substructure so that it is shaped and formed right, out of the right modular components.

We will need to pay particular attention to issues of intellectual property protection in the national research network environment, to help stimulate the creation of new creative works, and also to encourage people to make them available in the network environment.

Generally, I have discovered how difficult it is for the private sector to develop infrastructure on its own after trying very hard for many years to encourage industrial participation in this area. The potential for participation is there, but because these investments are of such a long-range nature, there is a very important federal leadership role that must be played in order to get the rest of the country behind this activity. I believe they will get behind it in time.

I also think it is very important to begin the development of this infrastructure with the research community because many of the ideas have yet to be developed. The research community can play a part in helping to create and develop them through the universities and research labs. Beyond that point, I think it can spread out to others who can then make use of what has developed. This is a critical area for federal support, I believe.

In my view, the network will have to be, a kind of evolving test bed. It should not be a static entity that, once developed, is then fixed at the end of a five-year period. I think there will be a need for continuing federal support for R&D on this network. As time goes on, parts of the network will become stable, and those parts ought to be prime candidates for transferring over to commercial services. The industrial sector ought to be capable of taking over the more stable parts as long as they are consistent with their own long-range interests in terms of both products and services.

It would be very desirable not to have someone be a computer science expert in order to be able to use this national research and education net. That is a very important reason why we need to develop these higher levels of the infrastructure so that the network is available to everyone who needs to use it and wants to use it and so that it can, in fact, gracefully extend to the rest of society.

In terms extending it into the educational community, I presume that we can all get behind the idea that we need to move beyond having this network just available to the top 50 or 100 colleges and universities in the United States.

I believe there will be a real utility in the network for the educational system at virtually every level. Furthermore, there is in my mind a clear utility to the rest of society as well. I think that we need to work with other parties in the United States in order to make that a reality.

Since education is the major responsibility of the states, getting the states involved in this process is a national priority and one that I think we have ignored to a large extent on the national scene. This is the key to the expansion in the educational sector, to the high school levels and beyond, and to the organizations that may not be able to compete for federal R&D funds or may not be even in the R&D business is the states.

I think the key to involving the rest of society in the network is to get industry to work with the states in furthering this adjective. The combination of the two will be essential in getting small businesses in the states involved. These are organizations that could use some local assistance. This will be an important step in ultimately getting the infrastructure to individuals in homes. Once we achieve this level of penetration, then we will truly have a national information infrastructure. We cannot have 1,000 or even 50 independent and uncoordinated activities and expect to end up with a national infrastructure. What we will have is 1,000 or 50 independent and uncoordinated activities.

In order to achieve a national information infrastructure we will need to create a new process of cooperation within the country not only in the test bed environment, but within all the participants that would like to participate. We need to develop the techniques and the protocols and the methodologies—partly these are social, partly these are technical—that enable the various activities to all play together as a coherent whole just as the road systems about at the state boundaries.

Finally, let me address what I believe the federal role in this whole process might be. I have some very specific recommendations but I wish to emphasize that this is just my own way of approaching the problem.

I very much like the highway analogy that you have talked about for many years. I really believe the network approval can be patterned after the highway bill. It has many of the attributes that we need for a national research and education net.

First of all, I think it is very important to provide sufficient federal R&D on a continuing basis for the network. This is not a onetime initiative—and after five years we walk away from it. Like our support for education, it must be a continual process that we take seriously.

Second, we need to involve the states directly in the process of building this infrastructure and in extending it to the educational communities at all levels. I think the states will be interested in this issue, not only for the purposes of education, but also for purposes of competitiveness and job creation within their states. We have had several discussions with representative from several states, and I believe there is wide interest across the country in that particular area.

Third, I believe that we ought to consider providing federal funds on a matching basis to match funds that are put up by industry and the states for infrastructure research and development. In the case of the highway system, the federal government put up 90 percent of the funds, and the states 10 percent, if I recall. There is more expertise in the government on this, perhaps, than elsewhere, but that is my understanding.

In the case of SEMATECH, which was an initiative begun by a group of industries in the semiconductor business, the federal matching funds were 1:1. In the case of infrastructure development, with a combination of state and industry money, a number like 3:1 might be an appropriate one.

Fourth, I believe we ought to consider setting up an infrastructure trust fund within the federal government to pay for the federal share of the operational costs of this infrastructure. There is a large amount of money that is going to be required to just run the network, and the government will have its share. Perhaps the helping trust fund could be adapted for this purpose.

The operational costs of the infrastrucure are probably too large for a single agency or department to absorb without skewing its budget presentation since most or all of the funding will be just strictly pass-through money. One could support the trust fund by appropriating a small percentage of the federal R&D budget each year as you mentioned earlier in your testimony. Perhaps less than 1 percent would do.

Five, even with a matching funds program, if everyone could then get the results without having to invest, there might be no incentive for industry to become involved. With the help of Congress, I believe that the SEMATECH funders were given access to its results for a period of about two years before it was made available to the public. A situation like that might be appropriate here, with some early access to encourage industry to fund some of the infrastructure, perhaps two or three years of early access before that portion is made widely available to everyone else.

Sixth, I believe that we ought to include federal government information in electronic form in the national information infrastructure. I think this could be an important adjunct to the private sector information, and I am thinking about not only outgoing information from the government, but also incoming information to the government.

Finally, I would like to emphasize again, because I think it is really a very important issue, that we ensure the protection of intellectual property rights in this network environment, both to stimulate the creation of new works and to encourage people to make them available in this environment.

Finally, I wish to point out the need to find a proper balance between the federal role, the states' role, and the role of the rest of the private sector in the actual creation of this national research network and the development of the infrastructure.

Thank you very much.

[The statement follows:]

TESTIMONY OF DR. ROBERT E. KAHN

PRESIDENT, CORPORATION FOR NATIONAL RESEARCH INITIATIVES

My name is Dr. Robert E. Kahn and I am President of the Corporation for National Research Initiatives (NRI), a not-for-profit organization founded in 1986 to foster research and development for a national Information Infrastructure. I am pleased to have this opportunity to testify before the Senate Subcommittee on Science, Technology and Space as you consider the role of a National Research and Education Network (NREN) as the backbone of a national "Information Infrastructure". At a hearing of this Subcommittee during the last session of Congress, at which I was also a witness, the occasion arose to touch briefly on the subject of a national Information Infrastructure. I welcome the opportunity to expand on this important subject.

I am very supportive of the efforts to create a NREN and to make it widely available for research and education. However, it is my belief that these efforts must go well beyond the development of the backbone network to include those enhancements which make the backbone into a widely usable infrastructure. I refer to these enhancements as the "higher levels" of the NREN which will make it a truly useful entity to all its potential users. By focusing on the research community initially, we can leverage the enormous talent in the country to help establish an evolutionary network-based infrastructure which can be used for research and experimentation, and can also be used for educational purposes. The results of this enterprise can subsequently be made available to the commercial sector for further development and use by the rest of society.

Federal funding for such infrastructure can be as valuable as any investment the government can make. The benefits will apply to every scientific discipline and can impact the productivity of the entire research and educational communities. I use the term "lower levels" of the Information Infrastructure to refer to certain supporting system capabilities (e.g., hardware, software, standards, protocols, system architecture) which form the backbone of the NREN. The higher levels of the infrastructure, such as the proposed Digital Library System, will build upon the lower levels to create specific systems that meet end user needs. Of particular interest are higher level systems that are difficult, if not impossible, for a single organization to develop by itself. Such systems might involve 1) use of networks for interaction among individuals and organizations and 2) common, industry-wide procedures to support and encourage widespread access. They can also be expected to facilitate common access to remote servers for information retrieval, for storage or for processing.

Information Infrastructure is not a single capability or system. It covers a totality of investments that make network-based information technology productive and useful. Thus, we should likely make the most progress by considering specific examples of infrastructure, rather than by treating the totality as if it were an entity by itself. Just as government and education are labels for many collective activities the totality of which defines them, so is infrastructure such a label.

Of particular importance to the NREN is its connection with education, the industrial sector, and those parts of society not directly involved with research. It is especially important that we address this connection since the national standard of living is profoundly tied to education and industrial productivity. In time, I have no doubt that industry (domestic or foreign) would provide some form of advanced infrastructure to support computer-based interactions. However, it is unlikely that an Information Infrastructure similar to the proposed gigabit NREN would materialize for decades based solely on private sector initiatives. I think it would be a mistake to wait for this to happen at its own pace. The bill under consideration today (S. 1067) is thus timely as well as important.

First, the economic potential derived from new and improved productivity tools is likely to be quite large. Second, the primary groups in the private sector with strong interests in this area cannot, by themselves, easily tackle the problems of long-range infrastructure development on a national scale. A major part of the problem is identifying the large sums of long-range funding for research and development that will be required to carry it out. Industry would need some assurance that a significant market exists (or will shortly exist) before making the needed investments. Without a federal initiative, the opportunities will remain just possibilities. Third, legal and regulatory limitations restrict what these groups can accomplish. Lack of a very high speed network infrastructure also limits what the computer companies can tackle. Thus, we have a classic "chicken and egg" syndrome. Fourth, not to be minimized, is the need to ensure the protection of rights in Intellectual Property in the NREN environment. The NREN should serve to stimulate the production of new creative works, such as databases, and encourage these works to be made available over the network. The protection of Intellectual Property rights will be a critical factor in achieving this objective.

NRI has been working with representatives of the publishers and the library communities as well as experts in the area of Intellectual Property to better understand this issue in the context of a national Digital Library System. Our tentative conclusion is that adequate protection exists under the combination of copyright law, contract law and the electronic communications privacy act to allow initial operation of the Digital Library System. However, further consideration of this area will be necessary as specific issues arise and we gain experience with the operation of the system.

NRI has also been working with members of the U.S. academic research community, with NSF and DARPA, and with industry to help foster the development of a gigabit networking capability in this country. I am confident that the technology can be made available for a gigabit NREN during the next decade. However, current estimates are that the equivalent costs per site for such a network will be at least an order of magnitude larger than for existing lower speed networks. It is widely believed that this increase in capability will be of enormous benefit to the research and educational communities once it becomes available as a useful tool. However, it will take a long-term commitment of additional government funding to sustain such a network. Some of these funds could be made available by the states, especially funds for education. In addition, industrial participation would initially help to offset costs. The assimilation of this gigabit networking capability into the commercial infrastructure of the country should enable it to become available to the rest of society at costs that are comparable to today's costs (after adjustment for inflation).

It is clear that gigabit networks will enable much larger quantities of information to flow at a given time. While it is possible to anticipate many of the benefits of such a network, I have not attempted to calculate the resulting value of such increased capacity for several reasons. One is that we are unlikely to know the real utility of such networks until several years after they have been available for actual use. Two, it is difficult to predict accurately the effects of large quantitative changes in technology that result in qualitative change. What seems most likely is that such networks will enable us to move from a text-based mode of interaction to one that is based on the movement and display of images. Thus, the user is likely to obtain technology which can support more natural modes of interaction through networks than is affordable at present.

Finally, the existence of gigabit networks will lead to many new and innovative products and services as well as the development of computers which can communicate directly with each other over the network at very high speeds. The implications of such developments are still wide open to conjecture.

The possibilities for research on networks and distributed computing systems are immense. However, an even larger opportunity exists to make the NREN a useful and productive tool for others as well. A scientific researcher should not be required to first become a network or computer systems expert to use the NREN. If that were a prerequisite, the cost of using the network, in human terms, would be too large. The NREN will become eminently practical for these users when the higher levels of the infrastructure are available and easy to use. One such capability is a Digital Library System that can be used for storing information in the network, for retrieving information from servers on the network, and many other purposes. By encouraging the development of educational material which is easily accessible or shareable over the network, the educational community can benefit greatly. Once the initial research is completed, the full impact of industrial contributions should begin to emerge.

While the NREN must be capable of continual evolution, it should be compatible with the long-range plans of industry. As state of the art technology and services are made available on a commercial basis, major portions of the NREN might begin to appear as standard commercial services. Yet, as new needs and requirements beyond the state of the art are identified by the research community, we must allow for ways to meet them. Just as today's fastest supercomputers will be tomorrow's desktop computers, so too can we expect access to today's fastest networks to be widely affordable at some time in the not-too-distant future. I believe industry will be very willing to coordinate this evolution with the government to insure a steady transfer of advanced technology and services from research into the commercial sector. Of course, I also expect there will be some capabilities that the government may determine it needs to pursue by itself, such as those which involve national security or other critical government functions.

Let me now turn my attention to the roles of industry and the states before addressing what the federal government might undertake in this context. Clearly, the academic community will benefit from its participation in the NREN, if government funding (federal, state and local) is made available for this purpose. What will it take for industry to participate in long-range infrastructure development? I believe it is essential for industry to be able to combine any altruistic objectives for an NREN with a sound long-range business rationale before widespread participation can occur. Industry has taken the lead in such areas on occasion, and should be interested in some level of participation in the NREN.

If the government provides adequate funds so that research and educational users can acquire advanced products and services from industry, there will be added incentive for industry to participate in the development of the technology and services. This, in turn, could spur industry to take a longer-term view of its investment strategy. In return, new state of the art capabilities could be available for early use in the NREN.

Five examples of major infrastructure developments which involved different combinations of participation by the federal government, the states and industry are briefly noted below. Numerous other examples could be cited, such as power systems, and rural electrification. In each of these cases, the benefits of the development extended far beyond the infrastructure itself.

a. Railroads - While most of the capital for the development of our railroads came from private sources, government was a major participant in the process. It either granted or facilitated the acquisition of land and rights of way. My impression is that the long term value of land along the rights of way was a major incentive to industry to participate in creating the railroads, rather than advanced technology or anticipated passenger and freight revenue per se. The land next to the railroad had particular value because it was next to the railroad.

b. Airlines - The federal government played a major role in fostering the development of this industry. The military interest in an air force was one early driver. Also the government was an early user of aircraft services for contract mail transportation. Cities and states participated in the creation of airports and access roads. The federal government created and still operates the air traffic control system.

c. Postal System - A federally created postal service was established across the nation with the use of private transportation facilities for long-haul carriage of mail. Industrial participation was minimal otherwise.

d. Telephone - Like the railroads, the telephone system evolved segment by segment, linked by standards and switching. Telephone calls were originally handled by many small local companies which sprang up across the country. Earlier in the century, the Bell system was assembled from many of these smaller companies. The Communications Act of 1934 and the Federal Communications Commission subsequently came on the scene to manage the growth and evolution of the communications and broadcast industries.

e. Highways - The justification for the highway system was based on military needs. However, the civilian implications were also quite obvious. The states played a key role in the overall process and paid ten percent of the development funds. A highway trust fund was established and paid for over time by gasoline taxes.

Turning now to the role of the states, I noted the important role they have played in many of the earlier infrastructure developments. I believe the states all have strong interests in technology development for reasons of economic competitiveness and the creation of jobs. The states should also be interested in assisting small businesses to get started and grow, and helping larger more established organizations to flourish. This help from the states may be essential, since many of these organizations may not be able to compete effectively for federal research and development funds, but could still benefit from state assistance such as partial support for a connection to the NREN.

More compelling, perhaps, is the fundamental role of the states in education. I know that much thought has been given to the role that computers and networks can play in education, but the full implications of an NREN have yet to be explored in conjunction with the states. I think this should be a national priority. Without state involvement, the NREN is unlikely to penetrate beyond the top echelon of schools, and, thus, most of the educational system would be excluded from this important development. It is not necessary that all fifty states decide to participate in the NREN. There will be ample time and opportunity to enter the process at any time in the future.

Once the NREN is implemented, the federal role could be patterned after aspects of the highway system. In addition to funding research and being a funding partner in the development of the infrastructure, the federal government could play a key policy role and help to form and maintain standards, in cooperation with the states and industry, as appropriate. In the long run, the best interests of the country will be served if we recognize, up front, the need to find a proper balance between the federal government, the states and the private sector in creating and maintaining this infrastructure.

My specific recommendations for the federal role are as follows:

1. Create the NREN as an Evolving Testbed

In order that the NREN can evolve and grow, I believe we must view it as a dynamic entity in which the well defined and stable parts would ultimately merge with standard commercial offerings, as new capabilities are developed to take their place. The research community needs a process which supports new ideas that can be experimentally evaluated. This will allow participants in the NREN to converge on agreed standards which arise as the result of experimentation instead of being artificially defined in advance of actual use. When we have sufficient experience using the testbed, the interested parties can take next steps to adopt and maintain national standards or to participate in international standards activities as appropriate. Over time, the NREN can be expected to evolve into the next generation experimental system that will continue to allow user needs to be identified and reflected in experimental systems before standards are finally adopted.

2. Provide Sufficient Funding for NREN R&D on a Continuing Basis

Many agencies of the federal government will be vitally interested in this aspect of the plan. While NSF and DARPA have played the lead role in the early gigabit research, DOE, NASA NIH, and many other government bodies should also be quite interested in participating. Efforts to coordinate federal research in networking has been undertaken by the Federal Research Internet Coordinating Committee (FRICC), and additional efforts will be required to coordinate the development of the higher levels of the infrastructure. Both the bill under discussion today (S. 1067) and The Federal High Performance Computing Program recently issued by the White House Office of Science and Technology Policy call for research support over a five year period. While it will surely be appropriate to reconsider the funding levels from year to year beyond that, I believe this funding should be viewed as the first phase of a continuing effort to grow and later evolve the NREN.

3. Create an Infrastructure Trust Fund

Once the NREN is established, it will take considerable resources to fund its operation. The federal portion of these funds should be placed in a separate trust fund to be allocated by the individual departments and agencies of the federal government based on research proposals and other similar criteria. It seems inappropriate to place these funds in the budget of any specific department or agency of the federal government as their mere presence can skew the perception of available funds within that organization. The NREN must have a stable identifiable source of support which is best handled by placing the funds in a separate pool. I would suggest yearly appropriations to the fund be based on an adjustable percentage of the total appropriated federal R&D budget for the year. An open issue which remains to be addressed is how to administer these funds within the federal government. Allocation of the funds could be based on a formula to be developed through FCCSET such as a proportional allocation to each department and agency based on its appropriated R&D budget (with the ability to transfer allocations between agencies). Or it could be based on the previous year's allocations.

4. Stimulate State and Industrial Research Funding

A portion of the trust fund should be set aside to "match" funds provided by industry and the states. In the highway system, the states put up ten percent of the funding and the federal government the remainder. In the case of SEMATECH, the ratio was 1:1. I would recommend a plan with 3:1 "matching funds" in this case. For every combined dollar made available for research by industry and the states, the federal government would put up three dollars, subject to available funding limits. These funds should be provided to states or private sector R&D organizations that 1) establish programs which are consistent with the development of a national Information Infrastructure, and 2) are also "qualified" by the government to conduct long range R&D programs on national Information Infrastructure. In addition to research and education, these programs would be expected to be broadly supportive of major segments of industry and the public. Clearly, this concept will require further refinement with the appropriate parties, but it could stimulate industry and the states to participate actively as funders of the long-term infrastructure.

5. Stimulate Industrial Contributions to Infrastructure Building

One way to stimulate industrial contributions is to enable early access to infrastructure for a period of a few years to those organizations that participate in funding it directly or via a matching grant program. This will encourage organizations to participate, since they may see a disadvantage in being left behind. Also, by allowing these organizations to participate in the planning process by which the infrastructure evolves, they can continue to affect its future trajectory.

6. Incorporate Federal Government Information in the National Information Infrastructure.

I realize that this is a sensitive issue with many unexplored dimensions -- but I have a feeling that the availability of federal information in electronic form could be an attraction for many participants, like the tract of land alongside the railroad. The question is how to make government information available electronically and on what basis? In recommending the incorporation of such information in the NREN, I am assuming that the government has the responsibility to make certain of its information available to the public and to ensure that it is available in some basic form at the nominal cost of distribution.

In the case of the proposed Digital Library System, I envision a two-tiered strategy for its implementation. The core part of the Digital Library System consists of information which is made available by the federal government (and most likely the states) and private sources, together with basic mechanisms for access. While government information would be available at the nominal cost of distribution, material made available by private sources would be subject to any limitations on allowed usage and additional charges for access to works protected under Intellectual Property laws. The second tier might convert that information into value-added products and services which would then be made available at additional cost. This strategy insures that users can still access government information at the nominal cost of delivery, privately owned material as the market dictates, and access to enhanced services at additional cost. We tend to think of infrastructure as if it were a monolithic entity. Just as infrastructure is not one thing but many, so is a particular piece of infrastructure not one homogeneous entity but rather many modular pieces produced by multiple organizations.

In summary, I am very supportive of the effort to create NREN. This will be an important capability for the research community and for improving our educational system. I have tried to emphasize the importance of developing the higher levels of the infrastructure and creating the mechanisms which insure it will grow and evolve. Industry and the states should be a part of the process from the beginning. This will ensure that the Information Infrastructure has a chance to permeate the entire fabric of society.

Senator GORE. Thank you very much. I appreciate all of our witnesses on this panel. I will have a few brief questions and then turn to my colleagues.

You used the word infrastructure, Dr. Kahn. Here on Capitol Hill, liberals and conservatives, Republicans and Democrats have always argued about the proper role of government in helping the country move into the future. There is a debate about trying to pick winners and losers. There are debates about all kinds of new government initiatives.

But one thing that all of the factions and groups and parties have usually been able to agree on is the wisdom of investing in infrastructure because a good infrastructure matched to the needs of our time can help everyone equally and can help anyone who cares to use it.

The problem is that this consensus about infrastructure was formed in an era when infrastructure meant water lines and bridges and railroads. We are now in the Information Age—and Dr. Billington very eloquently described the difference between the information age and the industrial age—and in the Information Age infrastructure must acquire a new, larger meaning. We must see information superhighways as part of a successful nation's essential infrastructure in order to compete in the future, and, as we have done in the past, help all of our citizens and businesses and schools equally to take advantage of the vast information resources that are now available.

I have used a simple example to describe the potential which I think we should strive for. Imagine a school child in Tennessee or South Dakota, Virginia or Nevada, just to pick some states at random, sitting at home with the equivalent of a Nintendo machine gaining access to the information in the Library of Congress.

My youngest child is fascinated by dinosaurs, and why should not a child who loves dinosaurs be able to explore through an interactive computer the information that is available? What is the largest dinosaur? What did it eat? Where did it live? How long ago was that? What does that mean? Did the earth look different then? Where were the continents? And, by means of natural curiosity, pursue knowledge to his heart's content.

This is one of the main ways we can improve the productivity of our educational system here in the United States.

Now, there are a lot of technical problems that have to be solved: copyright, design of the software, the protocols in the networks, and all of these things, but we can do it and we must do it.

Let me just ask all three of you: How realistic is it to dream of a child being able to interact with a library of information in that fashion? Can we do that?

Dr. BILLINGTON. Yes, we can, in fact, give you one example. The American Memory project which I indicated is entirely interactive.

In other words, you use optical disk, or some other variety of emerging technology, or perhaps several in combination, but you are in a sense presenting, on a disk or on line—there are various ways—a whole section of the library's collections which have been digitized, or captured electronically in some other way here.

But these materials are like a library itself—essentially useless, unless an active mind is interacting with it. You have to choose the books, you have to scan the materials, and one question leads to another, and, of course, the new electronic technologies enable you to stop where you want to.

Senator GORE. If you stop on a part of the subject of particular interest, you can—with the American Memory project—instantly scan through the photographs in the collection and look at what the people actually looked like. You can look at videotape.

Dr. BILLINGTON. Exactly. One aspect of American Memory that we think is really important is that it includes audio and visual material, as well as textual material. Also, it includes original documents, so that the student in a library, or in a school, or ultimately in his home at a personal computer, has it on the screen and is able to call up sections of the original document not just of text material but of the way the Founding Fathers wrote.

If you have the papers of Washington, Jefferson, Lincoln and so forth, as we do in the Library of Congress, it is wonderful to see—first of all, how good their handwriting was. You get all kinds of additional little bits of inspiration and challenge—how good the English was, how much they corrected themselves, how they edited themselves.

In a published version you just see little, dry anthologies, but in American Memory you can actually see the text. You can see the whole history of political cartoons, you know. You can see the cartoon of John Quincy Adams falling asleep when he was in the House of Representatives following his service as President. They gave him an awfully hard time.

You get a historical perspective on our problems. But the main point is, American Memory is interactive; and this approach has very practical results, not just in the study of history but in the sciences, where you can linger on a problem, open up areas of curiosity.

I think this is the wave of the future and the new technologies do permit it. We at the Library are moving in that direction and many others are, too. Of course, a national computer highway simply means that experiments that are successful in one part of the country can then be much more rapidly shared with another part.

Far from repressing the inventive genius of the American people and our inherent variety, it will open up much more variety.

The important thing, though, is the question of standards. With computerized information standards are even more important than with the book cataloging of the past, because computer protocols are precise. You are going to be using different kinds of hardware and that only increases the importance.

Senator GORE. That is another reason legislation is needed, to solve those problems.

I have already used up my time, but if the other witnesses wish to briefly respond to the same question.

Dr. KAHN. Yes. In any new technology, whether it is a digital library or a computer network, you can have simple things you do, more advanced things you do, or really forward-looking revolutionary things that you can do.

I am reminded of the days when we developed the ARPANET. There were many people who thought it was a straightforward thing to do—just put a bunch of telephone lines between different computers.

It turned out to be a far more complex situation than just putting telephone lines between computers—to actually think through all the manifestations of building what was at that time the very first, pioneering computer network.

We did the same thing in designing the INTERNET back in the early 1970s. The INTERNET is a collection of independent networks which were glued together through a system concept that we invented—a set of protocols to bring these networks together—a complex kind of activity that revived a common systems approach.

One could have done a simpler, more straightforward thing which would have been much less effective. The same is true of the digital library situation.

I am sure there are people who could argue that we had digital library systems back in the 1950s, or the earliest days of computers. In fact, you can have library systems today that are either very simple, or more advanced, or even revolutionary in concept.

My vision of a digital library system is one where the user does not have to know where the information is. The user can deal with multiple databases simultaneously, can ask simple queries and get back answers to complex retrievals. There is a tremendous amount of R&D that needs to be done to make that all happen, but that is my vision.

Mr. WYATT. A quick summary. Not only can we do it, as has been described here, but we must do it. The scenario that you presented in my view is the only way we are going to rapidly address the problem that we see, when our 13-year-olds finished dead last in international competition on math and science. That was pretty succinct and alarming.

Senator GORE. Senator Pressler?

OPENING STATEMENT BY SENATOR PRESSLER

Senator PRESSLER. Thank you, Mr. Chairman. I want to thank these witnesses. I think this hearing is one of the most exciting hearings that have occurred recently on Capitol Hill, because of the scope of it and what it means for our future.

Last Friday, the White House released a plan for a federal high peformance computing program. The plan embraces, with only a few minor differences, Senator Gore's proposal for a national research and education network. I support Senator Gore's bill, and I commend my colleague from Tennessee for his foresight and leadership in this area.

High performance computing and networks that link up users around the nation are essential to ensure the United States economic and technological leadership in the next century. The national research and education network proposed by Senator Gore has may applications that will improve the country's industrial competitiveness, advance our research on climate and weather, improve our ability to diagnose and cure diseases, improve our ability to predict natural disasters, strengthen our military readiness, and design and build the vehicles that will take us into space.

The network will provide an opportunity for researchers around the country to access databases in other locations, and to exchange information. It will especially benefit small and remote universities and research institutions that do not have access to supercomputer capabilities and specialized databases. For example, South Dakota State University is in the process of developing a biostress laboratory to develop strains of crops and livestock that are resistant to temperature extremes and drought conditions. The results of the research produced in the biostress lab could be placed on-line for ready access by users at other universities that are doing similar types of research.

Eventually, the network will be extended to include commercial enterprises. Then, small and medium sized companies will have access to the superior design and computational capabilities that large companies can afford today. For a small business, the savings that would be generated by using supercomputers could mean the difference between staying in business and succumbing to competition from larger companies and foreign competitors.

NASA's Earth Observing System (EOS) will soon be providing a large amount of global climate and environmental data for scientists to use in predicing climate changes. The land remote-sensing data will be stored at the EROS Data Center in Sioux Falls, South Dakota. A national network will permit instant access to this technology and data, and provide researchers with the ability to collaborate with scientists at other locations. This will significantly improve our ability to predict and react to changes in our climate and the environment.

The EROS Data Center in Sioux Falls archives the pictures that are sent back from our Landsat satellites. Since 1971, it has stored over one million satellite images on magnetic tape. That represents over 248 trillion bits of data. Government and university researchers and commercial enterprises use the images for a wide variety of applications, including oil and mineral exploration, land-use planning, mapmaking, monitoring environmental changes, national security purposes, determining insect damage in trees and crops, and predicting the size and health of crop yields. With the proposed network, any government, university, or private researcher could use the LANDSAT pictures that are stored in South Dakota without having to leave his desk or laboratory.

A computer network across the United States has great potential for increasing the productivity of American industry and stimulating the discovery and exchange of new ideas. I congratulate Senator Gore, and I look forward to hearing the witnesses today.

Let me say that I have listened with great interest, and I do have parochial interests, like all senators, but I am very concerned about the small businessman out away from a big city, and the smaller university.

We have the biostress lab at South Dakota University. It is not a big ten university. We do not have the resources that Vanderbilt and some other, larger universities have, but how can we be sure that this great source of information is going to be able to reach someone who is practicing medicine in a smaller town, or who is a small businessman?

Let me say, I think in terms of research and development, some of our cottage industries produce as much R&D as giant businesses do, and I am always amazed, going around my state, at the individual professors who have applied for grants from foundations or the government and are out there doing research virtually without staff support.

For example, there is some very good deep earth drilling research being done at South Dakota's School of Mines, which the Russians have done a great deal of and certain other countries have, but the United States does very little.

Now it is true, I suppose, of course, as you say, that telephone lines can hook things up and information can be there, but what steps can we take to be sure that this is not a Mississippi Valley University thing, an Ivy League thing?

We have the EPSCOR legislation that tries to get some of the research grants to the smaller institutions—and I know you cannot do all the research in smaller institutions. If you are doing research on Alzheimer's disease, it can get too scattered out, and so forth.

I guess I will start with Dr. Billington. You know, the great resources you have talked about, whether a professor is doing research on Shakespeare, or medicine, or whatever, as a practical matter, what do we have to do, or what has to be done, so that the person in smaller cities and towns and rural areas can access this?

Dr. BILLINGTON. I think we have to build networks. There are 110,000 libraries in this country. They are very widely scattered. Many of them are very small, but they can be on-line with libraries across the Country.

In the extensive review we have been doing in the Library of Congress trying to prepare it for the 21st Century, we have held more than a dozen forums in different parts of the country. I was out in the Plains states, and their problems are very different, but their library resources are quite substantial.

You have the school system, the educational system, with all its diversity, and the library system with all its diversity, and all you really need to do is to devise the kind of electronic highways that connect these, and in turn, with their local community and with the variety of Communities across the Country.

Now, for instance, we are going on line with our electronic cataloguing, which is an enormous searching device. We are trying it out with different kinds of libraries. In six months we are going to assess the results of this. What is the need, what is the utilization, and so forth?

I think in general the answer to your question is, there are existing networks. Ms. Avram is much more knowledgeable about networks—and I turn to her. But we use the existing networks, give them new resources, and they in turn will find the ways to relate to the local communities.

I do not think those are kinds of decisions that you have to make in Washington. You have to make the material available. We are, certainly, for instance, making our material available to different kinds of libraries—to highly scientific libraries, like Los Alamos, but also to some rural libraries in Kentucky, and so forth.

You advertize for applications, so that people come in with their own proposals, and then out of those you choose variety, do a testing to see how these networks become activated and then decide how you are going to do it. There are, I think, existing networks, and perhaps Ms. Avram would like to say another word about that, because she has been working with them for years.

Ms. AVRAM. In the library community there exists—it has been in operation for about four or five years now, under development for about ten—a methodology to exchange information with what we call the bibliographic utilities—OCLC, mentioned by Joe Wyatt, and that does some of what Bob Kahn was talking about, but it is only for the bibliographic data. It makes of the various databases a national database and allows the user not to know where the bibliographic record is. He just accesses information through this network.

Senator GORE. Sort of a Dewey digital system?

Ms. AVRAM. But I there are many things that have to be done, and of course one of them is—converting information in the future. We are finding more and more of it in electronic form, but if we are talking about our huge collections they must get into electronic form and that is rather a major part of it.

Mr. WYATT. Senator Pressler, may I add a dimension?

I think one of the things the proposed network will do is provide a great leveler for those people who participated in research around the United States in a single discipline and depend on access to either good data collections or good computer resources.

It is one thing to be standing in Harvard Yard and be able to walk into the Widener Library. It is quite another to not have that access.

If in fact we are able to achieve through the network and an infrastructure a system in which people can access material remotely, it will be the single most leveling effect on research and access to scholarly information and education that this country has ever seen. There is no doubt about that.

Then the game is essentially one of equal competition, and it puts it back on intellectual grounds, which I think is the place where we know it works best.

Senator PRESSLER. I think that is an excellent statement. You know, every year we have the same great debate in this committee and on the Senate floor about peer review groups, grants, research and allocation. My group, the people from my state, feels that the Mississippi Valley universities, the Ivy League universities, the California universities, the big ones, have a system to get all the research grants and the peer review guys, and it is a club.

So we passed EPSCOR, which requires a certain amount of research to go to small institutions. Of course, the argument goes both ways. Sometimes if you get good research too scattered out, it is not effective.

On the other hand, some of these small universities do very high quality work, and a lot of the R&D that happens in business comes from small business, not big business. So you can argue it endlessly, circuitously.

I do think if we set this up right and if it works out that people have access, it is going to be a renaissance of opportunity for people all over, wherever they are to participate, and it is going to be a great leveler. It will end some of those debates that we have. No, we will still have the debates, being Senators, but it will really have an impact on that whole argument, I think.

Dr. BILLINGTON. Senator Pressler, I might just say that we already answer hundreds of thousands of research inquiries that come disproportionately from communities that do not have access to the major research libraries. A system like this will enable you to do the same thing much more efficiently.

The example you mentioned of a mining school which would like to know what the Russians are doing or whoever is doing something competitively in the world, that is precisely the kind of information we have in the Library with the staff to read it.

A system like this or a better highway to carry the information from specialized locations would be precisely what would help your people. They do not want to build up a whole Russian scientific library. What they need is access to the ones that already exist. We need to get the material to the people that need it.

Senator PRESSLER. For example, how would the three or four professors who have a project on deep earth drilling at South Dakota School of Mines access the library if they wanted to know what is going on in the Soviet Union? Would it be in a visual form? It would probably be in a written form? Would they have to be hooked into the super computer, or what would the process be?

Dr. BILLINGTON. They can now make an inquiry, a direct telephone inquiry to the Library of Congress, or they can make a direct bibliographic search through one of the on-line libraries. I would have to see which would be the closest to South Dakota.

The kinds of systems that are envisioned here would make a direct electronic link-up and a rapid response that would include not merely bibliographic information and certainly already-stored information in data banks but, presumably, digitized information, full text information, scientific abstract information, etc. That is coming, I think.

It is expensive and would take time, but it is not as expensive as it might appear. We have even made some estimates what it would cost to do the Library of Congress collections. It is substantial, but it is not as awesome as I had expected it would be.

Senator PRESSLER. Mr. Chairman, I thank the witnesses.

Senator GORE. With Senator Robb's indulgence, if I may have 30 seconds to extend your last comment.

Joe Wyatt said earlier that the value of the Library to its users is substantially enhanced by the way in which they get new information and make it available. You talked about the expense of going back to the collect and digitizing it and making it available electronically.

It is important to note that where new information is concerned, most publishers now digitize the information before they translate the information into a typeface. It is not difficult to imagine the Library of Congress receiving new publications not only in printed form but directly over this network in digital form so that you do not have to then go back and retranslate it into digital form to make it available to users.

Senator Robb.

Senator ROBB. Thank you, Mr. Chairman. I should probably begin with a disclaimer that I frequently insert in any discussion of supercomputers and computer technology. I regret that I am one of those parents whose children are far more computer literate than they are. I constantly marvel at not only what they can do but the resources that are available to them. I have a great appreciation for the potential of the technology that we are discussing here this morning.

Although I have only a layman's understanding of this subject, let me ask a couple of questions that are in part technical. They might be better directed to the second panel, but let me just for my own edification see if I can get some response.

Late yesterday afternoon I was talking to an author/journalist who was working with on a PC. She had put a great deal of work into this particular story, and the computer somehow ate all of her story and research. She had done a backup tape and stuck that in. Again, I do not know how it was done technically, but the computer decided to eat that as well when she fed that in. She was left with nothing but her handwritten notes, and none of the intellectual work that she had done.

I mention this in connection with what we hear about computer viruses and other problems. If we have, in effect, an entire nation having access to all of this information stored in whatever form and in whatever place, is there a risk that someone could, either intentionally or otherwise, disable the system or cause information that has not been reduced to some written form to be erased or lost?

Mr. WYATT. Let me take a quick crack at this.

There are certain rules of thumb that one learns mostly through experience. One of the rules is never feed your backup tape to the same machine that ate the first one.

Senator ROBB. My guess is this particular person has learned that in a way that will not be forgotten.

Mr. WYATT. That is right. There are safeguards. In fact, many of the collections that we are describing here will be stored on materials that are not erasable; that is, on optical memories that in fact do not accept new information once they are recorded.

We have learned and are learning over a period of years to introduce a variety of safeguards that I think reduce danger to a minimum. One of the safeguards, quite frankly, is having backup storage. We are getting to the point where the cost effectiveness of this storage that we are able to use is so beneficial that we cannot afford to have just one backup.

Dr. BILLINGTON. The first rule made at the Library of Congress was to keep the old card catalog just in case.

Ms. AVRAM. We do exactly what Joe Wyatt said. We have several copies of the data base. We go as far as storing a copy outside the Library, because losing it would be a tremendous loss.

There is a danger of the computer virus. We have seen that. Of course, good heads are working on that problem. That can do a lot of damage.

Senator GORE. I might say that we have a section in the legislation that focuses on the virus issue and other security issues.

I agree with the Senator from Virginia. This is one of the absolutely critical things to focus on.

Dr. KAHN. I would simply add that while I think viruses and security in the network are important things to be continually concerned about, the more prominent case where a computer eats something is likely to be where there was no virus, there was no malfeasance involved but, rather, the user just did not know what to do to retrieve the data. It was there in the machine and with the right combination of esoteric commands and keystrokes it could have been retrieved.

It would be very nice if the higher levels of this infrastructure would let one say what happened to this data I just fed you and get it back.

Senator ROBB. I only used that to illustrate the concern on a broader level. In this particular case she took the computer and the tapes back to the manufacturer, and they could not find out what happened, either. So they are doing a little research.

Let me ask just one other question, if I may. You mentioned the rule of the states. I hope to encourage my own state to do something in this regard in the near term. We are looking at the possibility, for instance, of linking all of our major research universities with various technical databases and whatever the case may be.

My question is at what point do you achieve redundancy? When it is no longer useful, for a given university or jurisdiction, to continue compiling and making connections and then connecting into a larger system.

Is this something that really is not a grassroots, bottom-up system but ought to be begun and implemented entirely from the top down? Should we be encouraging states or other jurisdictions to hold up because they may soon be able to have access to the entire Library of Congress and all of these other major resources. Is there anything that you might do to make sure you would supplement existing files rather than simply create a redundant database or file?

Is that a problem right now? Should we concentrate all our energy at the highest level?

Dr. KAHN. I think you need both, Senator Robb. I believe that grassroots level initiatives are absolutely essential to get the right level of motivation and interest.

I have had discussions with various people in Virginia—at many of the colleges there, the Center for Innovative Technology, they are all very interested in this, but unless you have the grassroots it does not really take.

But the grassroots alone is not sufficient. You really do need a more global view, because whatever Virginia does you want to make sure is compatible with Maryland and Tennessee and South Dakota and whatever states are involved. That is a top level issue.

Senator ROBB. If we are about to spend, say, $15 to $20 million on the start-up costs for our Center for Innovative Technology, how could we be certain that we are not spending a great deal of that money in ways that will mesh with the national system? We don't want to waste money, in effect, at this stage when we do not have this superhighway fully in place?

I am trying to get some sense of what to avoid in the early stages so that we can use scarce resources as productively as possible.

Ms. AVRAM. I believe that is the infrastructure we are talking about. Part of that is a very important element of this highway that we are building, and that is called standardization, and that has to be coordinated and everybody has to adhere when that happens.

One of the tremendous advantages of the network will be eventually—not right away, but eventually—that the network will remove much duplication and thus it will be economically advantageous.

Dr. KAHN. I actually think you are 90 percent of the way there by just knowing that is a problem up front. This is not a formula issue. I believe it is one of sitting down with all the parties, identifying what needs to be dealt with on a global basis, what can be dealt with on a local basis, and hammering it out.

There will be some cases where low-level decisions about how many circuits and where to place them can be done on a basis of computer programs that are well engineered to design networks or probability and statistical analysis about probabilities of failure. That is a cost tradeoff involuntary of how much are you willing to pay for how much reliability.

But most of the other issues you are talking about, ought to be resolved in the process of discussion among all the involved parties. They can never decide at what level attention needs to be given to a given problem.

Senator ROBB. There are any number of laws passed by the Congress, where standardization, is a problem. Is it fair to say that each individual supercomputer, for instance, that might be part of this system is going to have a value that is not dependent upon how it is hooked up with the others? Or will it retain its individual value only for certain users who are a part of a particular network? If there are dozens of incompatible networks, then you are simply providing, again, redundancy or overlap that you could avoid if everybody got together.

In the best of all worlds, that is the way we would do a lot of things, but we do not. Everybody moves at their own speed and, regrettably, the Congress does not always anticipate all of the state and local and other jurisdictional needs or initiatives, so we have to find a way to mesh them later on.

I am just trying to get some sense.

Ms. AVRAM. I think what we are building here, we are building on a group of systems that are not necessarily alike and probably will never be. We would have to go back in time and undo everything we have done to make everybody operate their computer in exactly the same way.

What we must do is make sure we have standardization at the point of interchange, at the point of communication.

Senator GORE. But his question is about an institution in Virginia that purchases a supercomputer. It is not going to make making a mistake by doing so if it does not immediately link up to the network? I mean, the supercompuer has inherent value which will then be magnified once it gains access to the network? That is correct, is it not?

Ms. AVRAM. I would say the supercomputer put in in Virginia is going to have to add whatever it takes to become part of the network. In the cases that Bob was describing there was a language, a protocol, and each machine was able to communicate with every other machine using that standard language.

Senator ROBB. Just one final parochial pitch, if I may. In collocated with our Center for Innovative Technology that we have the Software Productivity Consortium which is trying to make sure that all of these systems can talk to each other worldwide so we may be able to solve this standardization problem at some point down the road.

Mr. Chairman, I thank you.

Senator GORE. Thank you very much, Senator Robb. We had many questions for this panel that we will not get to.

There are a lot of other questions, but we have a lot of expertise on the next two panels, and we are going to have to move on. But, you four have gotten us off to a great start, and you leave with our gratitude. Thank you very, very much.

Our second panel is made up of Dr. John Fischer, accompanied by Dr. David Nystrom, and Dr. Daniel Masys. If the three of you would come up to the witness table.

Dr. John Fischer is Acting Associate Director of the United States Geological Survey. Dr. Nystrom, who accompanies him, is Chief of the Office of Geographic and Cartographic Research at the Geological Survey in Reston, Virginia, and Dr. Daniel Masys is Director of the Lister Hill National Center for Biomedical Communications at the National Library of Medicine in Bethesda, Maryland.

We appreciate you three being here. As with the previous panel, your prepared statements will be included in full in the record. We would ask you to summarize the important points you wish to make, and we will begin, Dr. Fischer, with you. Welcome, and please proceed.

STATEMENT OF DR. JOHN N. FISCHER, ACTING ASSOCIATE DIRECTOR, U.S. GEOLOGICAL SURVEY; ACCOMPANIED BY DAVID A. NYSTROM, CHIEF, OFFICE OF GEOGRAPHIC AND CARTOGRAPHIC RESEARCH; AND LAWRENCE BATTEN

Dr. FISCHER. Thank you, Mr. Chairman. I appreciate this opportunity to testify before you today. My testimony will focus on activities of the USGS in data management and touch briefly on our role in the coordination of data management across the government.

I also will comment on our use of computers, computer networking and Geographic Information Systems. Optimal use of Geographic Information Systems relies on efficient data management, an objective of S. 1067. Therefore, at the conclusion of my statement David Nystrom and Lawrence Batten, who are accompanying me today, will give a brief demonstration of GIS technology and LANDSAT imagery.

Mr. Chairman, the USGS is deeply interested in the availability of supercomputers and networks to utilize them. The successful accomplishment of our mission requires that we apply appropriate computer technology to the collection, analysis, and dissemination of earth-science data.

Our use of computers spans the spectrum of technology from microcomputers to work stations to mainframes and supercomputers. We serve a variety of users in the public and private sector, each of whom have different capabilities to process the data they receive.

With the proliferation of GIS software and methods for collecting and representing spatial information, it has become apparent that the exchange of data between scientists and users of scientific data is a significant problem.

In response, the Department of the Interior has established the Interior Digital Cartography Coordinating Committee and OMB has established the Federal Interagency Coordinating Committee on Digital Cartography. These committees are charged to recommend procedures and programs, including data standards and data and exchange formats, to facilitate the coordination of digital cartographic data in the Federal community. The USGS chairs both coordinating committees and USGS scientists are actively involved in GIS activities and work groups of the committees.

The rapid growth of data derived from computer modeling and satellite sensors such as LANDSAT has resulted in an increased need for and a new focus on management of these data throughout the Federal Government.

The USGS data center in Sioux Falls, South Dakota, is deeply involved in the research and implementation of information management systems for image data. We are using satellite image data from Sioux Falls for a wide variety of applications ranging from large-scale wildfire monitoring to mapping drought conditions. We expect the USGS data center to play a central role in archiving and distributing data associated with global change problems.

At the conclusion of my testimony, Dave Nystrom will provide one example of a GIS wildfire monitoring application using LANDSAT imagery of Yellowstone National Park.

Microcomputers and minicomputers are an integral part of our scientific equipment in over 250 USGS offices across the country. These computers, along with the mainframe in Reston, are connected by a network called GEONET which allows users to connect to any other computer in the network. However, and now related more specifically to S. 1067, it is clear that solutions to some of the problems on which we are working would be facilitated by faster networks and more powerful computers than are available to us currently.

As just one example, scientific analyses of our changing planet are complex and multi-disciplinary. The interactions between the solid earth, the atmosphere, the biosphere, and the human influence on these processes are central to understanding the important changes taking place around us.

These interactions require geologists and geophysicists and hydrologists to share data bases and knowledge with biologists, agronomists, meteorologists, economists, demographers and scientists from a multitude of other disciplines.

Increasingly in the process of these multi-disciplinary analyses, very large and complex data files are created, moved about, and manipulated by the many different organizations and sciences involved in the activities.

Advanced technology computers and the high-capacity networks to serve them will be indispensable components in handling the massive volumes of data and the complexity of multi-disciplinary process modeling that will be an important part to successful analyses of global change.

Mr. Chairman, this concludes my statement. I thank you again for the opportunity to testify before you today.

[The statement and questions and answers follow:]

Testimony of Dr. John N. Fischer,
Acting Associate Director, U.S. Geological Survey

Mr. Chairman, Members of the Committee, I appreciate this opportunity to testify before you today and to describe some elements of U.S. Geological Survey (USGS) activities in data management.

My testimony will focus on data base development and management in the USGS, and I will comment on our current use of computers and computer networking, other new information dissemination technologies, and our current and future use of geographic information systems (GIS) in the USGS. At the conclusion, David Nystrom and Lawrence Batten who are accompanying me today, will present a brief demonstration of current GIS applications.

The USGS is deeply involved in the application of computer technology to the collection, analysis, and dissemination of earth-science data. Our use of computers spans the spectrum of technology, from microcomputers to workstations to mainframes and supercomputers. We serve a variety of users in the public and private sector, each of whom have different capabilities to process the data they receive. Some of our books and maps now are available not only on paper, but also on magnetic tape, optical media, and in certain cases, by direct access to computer-stored data. We are moving to an environment of digital data with dissemination by CD-ROM (Computer Disc-Read Only Memory) as a distribution medium, and by direct computer network access to our major data bases.

Large amounts of data are being digitized at numerous locations throughout the country by, or at the request of, local and State agencies and other Federal departments. With the proliferation of GIS software and methods for inputting and representing spatial information, the exchange of data from one system to another has become an increasingly difficult problem. Standard digitizing specifications and recommended procedures are being established and implemented by the Geological Survey to bring uniformity and standardization to the data. In addition, methods for transferring data among various earth-science data bases and available GIS's are being developed.

To address the extensive national needs for up-to-date digital earth-science data for GIS applications, the USGS is expanding significantly its digital data production capabilities through Advanced Cartographic Systems, developed in conjunction with the Defense Mapping Agency. These activities are currently focused on accelerating the generation of digital cartographic products, incorporating digital revision techniques, and integrating the new technology for automated extraction of data, image processing, and product generation.

In the early 1980's, with the proliferation of GIS software and methods for collecting and representing spatial information, the exchange of data between systems was a difficult problem. In 1982, the Department of the Interior established the Interior Digital Cartography Coordinating Committee. In 1983, the Federal Interagency Coordinating Committee on Digital Cartography was established by the Office of Management and Budget. These committees were given a mandate to recommend procedures and programs, including data standards and exchange formats, to facilitate the coordination of digital cartographic data in the Federal community. The USGS chairs both coordinating committees, and USGS scientists are actively involved in the GIS activities and work groups of the committees. A number of reports and GIS documents were recently published by the committees and are available from the USGS.

The rapid growth of data derived from computer modeling and satellite sensors (LANDSAT data, for example) has resulted in an increased need for, and a new focus on, the management of these data throughout the Federal Government. Scientists require improved access to the data as well as greater distribution of data products. In response to this need, the Geological Survey is coordinating with other scientific agencies in several collaborative endeavors to improve sharing and access to this data. Four of these endeavors are: the establishment of an earth-science data directory, the development of an Arctic environmental data directory, participation in a global change data and information system, and the establishment of a communications network called GEONET.

In 1985, the USGS established the Earth Science Data Directory (ESDD), which contains references to earth-science and natural resource data holdings of the USGS, other Federal agencies, and State agencies in all 50 states. The ESDD is an integral part of the shared directory activities of the Interagency

Working Group on Data Management for Global Change (IWG). Many of the data sources referenced in ESDD offer potential leads to data that might be used in GIS applications.

Several years ago, when the USGS and other scientific agencies formed the IWG, its objective was to make it as easy as possible for the scientific community to access and use data needed to study global change. A goal of the IWG is to develop a global change data and information system to be available by 1995 that is consistent across agencies and involves and supports university and other user communities. The system will use existing data resources to the maximum practical extent to make access to global change data as easy and inexpensive as possible for the research community. The IWG is coordinating its efforts domestically with the Federal Coordinating Council for Science, Engineering and Technology's Committee on Earth Sciences, which is responsible for developing the U.S. Global Change Research Plan with the National Academy of Science's Committee on Geophysical Data, and internationally with the International Geosphere and Biosphere Program.

During the past year, the USGS has cooperated with the Arctic Research Commission and the Interagency Arctic Research Policy Commission to coordinate efforts within the United States to make Arctic environmental data more accessible to Arctic researchers. As part of these collaborative efforts, the USGS established the Arctic Environmental Data Directory as a subset of the ESDD to store references to Arctic data holdings from governmental, academic, and private sources. Because Arctic processes are important to the understanding of global climate change, the USGS is cooperating with other scientific agencies and other Arctic nations to share the data of the Arctic community with, and connect its resources more closely to, the global change program.

We have installed microcomputers and minicomputers in most of our 100 major offices. These computers, along with a mainframe computer in Reston, are connected together by a network called GEONET, which allows users to connect to any computer on the network. The network also allows smaller offices that don't have their own computer to connect to the nearest office that has one. A subset of this computer network is called the Distributed Information System, which has approximately 55 minicomputers and provides a wide variety of textual, numeric, graphic, and image processing.

In Fiscal Year 1990, it is expected that a new set of workstation computers will begin to replace our existing minicomputers. Under the planned enhancements, each office will have several powerful workstations, delivering computer resources to the end user more directly and providing additional resources to match increased demand. The ability of USGS scientists to share data sets and software across the present GEONET has already proven to be a major factor in the spread of GIS activities within USGS and among our State and Federal cooperators. Since GIS data sets tend to be larger than most scientific data sets, a high-capacity network would be of immense help in allowing scientists to share even larger GIS data sets expected in the future.

Recent advances in GIS capabilities are bringing about an increased efficiency in acquiring, storing, manipulating, analyzing, displaying, and disseminating earth-science data that was not possible 2 years ago. While the operational programs of the USGS that apply GIS technologies are highly distributed, the GIS research and development work is more centralized. The USGS is developing a coordinated program of GIS research and development that will be necessary to successfully accomplish increasingly complex program mandates. Many current earth-science issues, such as landslides, earthquakes, contaminated ground-water systems, and global change phenomena, are problems that GIS technological advances now provide a unique opportunity to research, analyze, and solve.

The USGS has consolidated selected GIS hardware/software and personnel into shared regional research and development facilities at the USGS National Center at Reston, Virginia, and in Denver, Colorado, and Menlo Park, California. The consolidation makes available, at key regional locations, a full range of computer-based hardware and software. These centralized facilities, available to both the USGS and other Federal agencies, have been especially helpful in starting pilot projects, designing new data bases, and making decisions about appropriate equipment and software. Additional regional GIS capabilities are available at USGS facilities in Sioux Falls, South Dakota, Anchorage, Alaska, Rolla, Missouri, and Stennis Space Center, Mississippi.

We believe that the consolidation of USGS hardware, software, and personnel into shared research applications and development laboratories will continue to provide a powerful, cost-effective tool for USGS scientists to analyze and interpret data about land, water, and mineral resources to meet programmatic responsibilities. The research and development activities of our laboratories provide the technical basis for the methods for representing spatial information and exchanging digital data from one system to another. Results from a variety of cooperative applications projects are enabling USGS to make appropriate responses to digital earth-science data user needs at the local, State, national, and international level.

This concludes my prepared statement. We would now like to present a demonstration of our current GIS applications.

QUESTIONS OF SENATOR GORE AND THE ANSWERS THERETO

Data Management

I have been in the Congress long enough to know that when someone mentions "data management" a lot of people on the Hill just stop listening. Data isn't something you can put your hands on, and rooms full of magnetic tape don't look very exciting. The result is that data management is often underfunded. Sometimes an experiment is run, the data is collected, but then there's not enough money to store the data properly. As a result, the data goes unused.

How can we make people recognize the critical importance of data management?

We should promote the idea that good data management is essential for good decisions. Different methods will be effective in promoting this idea with different groups of people. With policy makers, good national policy decisions are based on a good understanding of available information. Policy makers must first be aware of the existence of the appropriate information and then have easy access to it. For example, the increased use of information from Geographic Information Systems (GIS) by policy makers has generated a tremendous interest in high quality data sources for decision-making. Many users of GIS, however, have discovered that it is difficult to identify sources of well managed accurately positioned data. To encourage scientists to manage and preserve their data, Federal agencies must include data management as a formalized part of conducting scientific investigations by planning to manage their data well and by providing the necessary funds to perform data management activities.

Computers now allow researchers to collect much more data than they could in the past. NASA is now working on EOS, the Earth Observing System, which will produce trillions of bits of data a second. Will the USGS and other agencies be ready for this flood of data?

The USGS intends to be prepared for the data transmitted by EOS. In March 1988, NASA and the USGS signed a Memorandum of Understanding for Experimental Land Remotely Sensed Data Processing, Distribution, Archiving and Related Science Support. The USGS EROS Data Center will serve as the archive for experimental land remotely sensed data obtained by NASA. The USGS is participating with NASA in developing and implementing the EOS Data and Information System (EOSDIS) for the mid-1990's. In concert with the Interagency Working Group on Data Management for Global Change, the USGS also plans to develop and operate global change land data management and information systems and will continue to operate the National Satellite Land Remote Sensing Data Archive.

What kinds of technology would be most useful for the USGS and the EROS data center?

The USGS will need advances in data storage devices, telecommunications, computer hardware, and computer software. Improved digital data storage devices will be required for efficient, reliable storage and retrieval of large data sets. High data transfer rates are needed to move large data sets from the storage media to users' computers. Supercomputing technologies need to be applied to data management to allow development of advanced algorithms for data correction and enhancement and more rapid movement of data between computer peripherals. Software improvements are also needed, such as high performance relational systems, more efficient data compression techniques, and advances in object-oriented data base management systems.

Saving the Data

Over the years, taxpayers have spent billions acquiring scientific data. One of the primary goals for S. 1067 is to see that all that data is accessible and put to the best possible use. I would really appreciate your honest opinion on this simple question:

Has the Federal Government done enough to preserve all the data it's paid for over the years?

Probably not. Although many cases exist where data have been preserved in an exemplary manner, cases have also occurred where data have disappeared or disintegrated because of inadequate data management.

Where are the biggest problems?

Federal scientific agencies do not have a consistent definition and understanding of data management, and adequate data management practices are often not established or enforced for research scientists. Documentation of scientific data is a persistent problem. The existence and content of valuable research data is often not documented by scientists and, thus, the data is not readily available or useful to others. Many agencies cannot clearly define much of the data they have developed and how that data can be accessed. Because of the large volume of data being collected and generated, maintaining the continuity and quality of the data is a great challenge.

I imagine that some agencies do better than others in this regard. Which of the science agencies are doing a particularly good job?

In general, the Federal agencies with a scientific data management mission do a good job in managing their data, and centers of data management excellence exist within each of the science agencies. These centers provide for scientists working with data managers to oversee the valuable data resources. The USGS is the principal Federal agency for the earth sciences and maintains many formal, large-scale geologic, hydrologic, and cartographic data bases systems and information clearinghouse services.

What else needs to be done?

A widely-accepted and consistent data management process should be adopted by the science agencies. Data management systems must be realistically costed out on a life-cycle basis, developed using the most appropriate technology, competently managed, and given appropriate priority. An important early part of the data management life cycle is determining what data will be collected, what will be saved, what data will be on-line, and what will be archived. It is not realistic to retain every bit of data that is collected. Just as information has been managed in a consistent way by a network of research libraries, data must become an integral part of our scientific processes. Benefit/cost analyses for multiple-purpose data systems must take into account out-year benefits.

Data exchange standards will help to facilitate data access and use, and significant progress is being made in implementing such a standard for spatially referenced digital data, which has broad applicability to geographical analysis, the earth sciences and biological, socio-economic and other studies. Because of the volume of data being collected and generated, maintaining the continuity and quality of the data is a great challenge.

NSF's Role in the Information Infrastructure

The legislation we are considering would give the National Science Foundation responsibility for coordinating the linking of scientific databases at different Federal agencies to the 3-gigabit network. Since NSF is running the network, it makes sense that it would have that role.

Which of the data bases at the USGS do you think will be hooked up to this national network? How hard do you think that will be?

The USGS maintains many extremely valuable data bases that are widely used by the scientific community and the public. Some contain scientific data while others contain indexes to data and information managed by the USGS and other organizations. Included among the USGS data bases that could be linked to the national science network are the National Digital Cartographic Data Base, Geographic Names Information System, National Satellite Land Remote Sensing Data Archive, National Water Information System (previously WATSTORE), National Water Data Exchange, Water Resources Scientific Information Center literature abstract data base, National Coal Resources Data System, Mineral Resources Data System, Rock Analysis Storage System, INTERMAGNET (joint international system for collecting geomagnetic data), quick earthquake epicenter data, earthquake data base, on-line catalog of USGS library holdings acquired since 1975, Geologic Long-Range Inclined Asdic data, National Uranium Resource Evaluation, Side-Looking Airborne Radar, and aircraft scanner data holdings.

The USGS has been using data telecommunications for many years. In fact, most of the data bases mentioned above are already accessible through GEONET, the data telecommunications network the USGS operates for the Department of the Interior (DOI). GEONET is connected to the NSF supercomputer network and other national networks. Some long-term means such as GEONET must be provided to connect USGS and DOI computers to the proposed national science telecommunications network. The hardest part of connecting data bases to a network is not the telecommunications links, however. It is agreement among network partners on consistent data directories, catalogs, user registration and access procedures, and the user interface--that is, the presentation of information and choices to users of the data bases.

Do you think that NSF is the logical candidate for coordinating the networking of the different data bases we are talking about? What kind of coordination would you like to see NSF provide?

We do not believe that there is a logical single candidate to coordinate the networking of the different data bases that would be linked by the network. This is a discipline-oriented function requiring significant input from each of the major science and engineering specialties. There are established coordinating bodies and mechanisms to help with the networking. Examples are the Federal Interagency Coordinating Committee on Digital Cartography, Committee on Earth Sciences, and the Interagency Working Group on Data Management for Global Change. In addition, OMB Circular A-67 assigns responsibility for Water Data Coordination for all Federal Agencies to the U.S. Geological Survey. Because of the demonstrated success of the NSF supercomputer network, we believe it would be logical for NSF to have the same type of research and development support role for the proposed broadband national supercomputer and science information network.

Coordinating Federal Data Management

Literally dozens of Federal agencies are involved in computer research, computer networking, and data management. One of the goals of S. 1067 is to provide for better coordination between the agencies.

Good coordination is going to become more important in the future because data rates are increasing so fast. The Earth Observing System, scheduled for launch in 1996 is expected to generate one terabit of data each day. That's a trillion bits of data. Every few weeks the EOS spacecraft will generate as much data as is currently stored in all the books in the Library of Congress.

It will be hard for agencies to share this data because federal data banks and computer networks grew on an ad-hoc basis and many of them are incompatible, making it more difficult to exchange data between agencies.

What can be done to improve coordination of Federal computer systems and communications networks?

The National Institute of Standards and Technology (NIST) should continue to work toward the establishment of new Federal Information Processing Standards (FIPS) for networking and software portability and continue the refinement of existing standards in those areas, such as Government Open Systems Interconnection Profile (GOSIP) and Portable Operating System Interface for Computer Environments (POSIX). In addition, the Federal government should capitalize on the successful experience of national data centers such as the National Climate Data Center, the EROS Data Center, and the National Center for Atmospheric Research, and successful network operations such as NSFNET and GEONET. Finally, we should build on the good models that now exist for interagency coordination of complex, joint technological ventures. Such a group could be empowered to coordinate the linkage of scientific data sets residing in Federal agencies.

What is most important?

The Federal science agencies must become active participants in the coordination of data management and networking for the linkage of and access to Federal scientific data.

What are the principal issues in ensuring that users can effectively and conveniently access and integrate information from a variety of data bases and a variety of networks?

It is essential for NIST to continue to develop and refine data standards such as POSIX and GOSIP. In addition, standards for data-set documentation, data exchange, storage technology, and data management should be addressed by an appropriate coordinating committee. The agencies must then adopt and implement the standards that are established. Training will be required to raise the skill levels of data management system developers and potential users. In addition, distributed operating systems and distributed data base management software will be needed to link multi-vendor and multi-generation computer systems.

The USGS is lead agency for computerized mapping. Are the agencies working together well in that area?

Yes. Substantial progress is being made, especially in the areas of standards and data exchange formats. Under the auspices of the Federal Interagency Coordinating Committee for Digital Cartography and associated professional, academic, and industry groups, new standards for Digital Line Graph (DLG-Enhanced) and Spatial Data Transfer (STDS) are undergoing evaluation. The agencies have a strong incentive to cooperate because advances in geographic information systems (GIS) technology offer an enormous potential for planning and analysis and make data exchange an economic imperative.

Picking the Right Technology

Advances in computer and communications technologies hold a promise for providing solutions to many of the nation's federal data management problems. However, these rapid advances also introduce the possibility that existing systems will become obsolete much sooner. For example, many agencies have data that was recorded ten years ago and that cannot be used today because the machine that recorded it is no longer available.

I understand that this has been a problem at the EROS Data Center. How is the USGS planning for future advances in data management technology?

With the encouragement and technical support of the EROS Data Center (EDC), NASA started to convert Landsat Multispectral Scanner (MSS) images collected from 1972 to 1978 to computer-readable magnetic tapes. For various reasons, the conversion was not completed and EDC has inherited the wide-band video station tapes (WBVT) and an obsolete processing conversion system. Over the next 4-5 years, as part of its National Archive responsibility, the USGS will try to finish the conversions of post-1978 MSS and Thematic Mapper data to a next-generation storage media.

To prepare for future advances in data management technology, the USGS is implementing FIPS networking and portability standards and is coordinating the testing of the proposed Spatial Data Transfer Specification. The USGS is also prototyping and evaluating evolving technologies. In addition, the USGS must develop a prudent mechanism for determining what data will be collected and what will be saved, what data will be on-line and what will reside on archival storage. Technology is not always the answer; rather, minimizing the problem first may allow the USGS to keep pace with technology.

What criteria should be developed in the selection and acquisition of computer and data storage equipment to ensure that archived data is readable in the future?

Federal agencies must agree on strategies for the migration of data across technologies. We must accept the fact that long-term digital data archives using magnetic media still have to be re-recorded on a 10-15 year (or shorter) cycle. This approach can be effective only if the archive management organization is mandated to carry out long term data retention, and therefore has the commensurate funding and management responsibility to insure that data conversion to new media is planned for and carried out.

Are the emerging computer and data storage technologies robust enough to ensure that the Federal Government will be able to manage the expected volumes of data?

The weak link in data management is computer software, including data base management systems and directory interchange software. Computer and data storage technologies are advancing and have generally kept pace with the evolution of data systems and increased data volumes. However, it is not realistic to expect all data storage problems to be solved by fortuitous technological advances. Responsible decisions must be made to identify those data that should be retained and those data that must be accessible interactively versus those that must be accessible within a reasonable time period. Alternative distribution media such as optical disk should also be considered for data that requires widespread distribution.

Supercomputers can generate and process hundreds of millions of bits of data a second, and they are getting faster. Are data storage systems going to be able to keep up?

Data storage technology will likely be adequate to meet the demands of supercomputer users. Because supercomputers generate large volumes of data, however, decisions must be made, based on economics and frequency of need, on whether to retain new data created by the supercomputers or instead, regenerate the data when it is needed again. In addition, it is essential that users have a means of visualizing supercomputer results to facilitate rapid analysis.

What can be done to reduce this gap and what role should the Federal Government play in this effort?

Federal agencies could acquire and use experimental hardware and software and conduct research to take advantage of new technology. This would provide an incentive to the computing industry to develop innovative products and to the science community to learn the relative merits of new technology in the pursuit of knowledge.

Who Uses USGS Data Bases?

The USGS maintains a number of different databases, including Landsat data and many different types of map data.

What other data bases do you have at USGS? Which ones are the largest?

The largest data bases include the National Digital Cartographic Data Base, the Landsat data collection, National Water Information System, and a major data base of analog aerial photography from the National High Altitude Photography and National Aerial Photography Programs. Other large data bases include the Geographic Names Information System, the Advanced Very High Resolution Radiometer sensor data, Water Resources Scientific Information Center literature abstracts data base, National Coal Resources Data System, Mineral Resources Data System, Rock Analysis Storage System, INTERMAGNET (joint international system for collecting geomagnetic data), quick earthquake, epicenter data, earthquake data base, on-line catalog of USGS library holdings, Geologic Long-Range Inclined Asdic data, digital data from the National Uranium Resource Evaluation program, and Side-Looking Airborne Radar data.

Who are the primary users of the USGS data bases?

The major users of USGS data are Federal agencies and State and local governments. The USGS, itself, uses the data in direct support of its mission. Other users include private citizens, academic institutions, foreign governments, and private and nonprofit organizations. Numerous requests for USGS data bases are routinely received from private data dissemination organizations which repackage and market the data.

Are your data bases used by just a few researchers in a few specific fields or are they broadly used?

USGS data bases are broadly used by literally thousands of private and public organizations and for many different purposes. For example, each year the EROS Data Center receives over 50,000 inquiries and orders for remotely sensed data of specific sites. Annually, the USGS answers about 85,000 requests for water data from government agencies, private consultants, the academic community, and the general public. Computer links exist between the USGS and more than 400 Federal, state and university data bases. In addition, several key USGS data bases are released periodically on Compact Disc Read-only Memory (CD ROM) for low-cost use in the public and private sectors.

How useful would it be for the network to provide access to them? What would be the main advantages?

Network access would probably be used most by Federal, State, and local governments, and larger private minerals development, environmental, transportation, and engineering interests, and selectively by academic institutions. Online access, however, is not always appropriate for many of the data holdings. For example, the collection of data on a given geographic area or on selected themes might be more cost-effectively distributed on optical media. There must be a tradeoff between ease and frequency of access and the quantity of data.

What kinds of network speeds are necessary to use networks to distribute Landsat images and other remote-sensing data? Megabits per second? Gigabits?

To distribute Landsat images and other remotely sensed data, transmission speeds in the high megabit range are required at a minimum; however, it has not yet been determined whether wide-spread telecommunications access to remotely sensed data would be cost-effective or even necessary.

What are the main barriers to providing for access to USGS data over networks?

The main barriers to providing access to USGS data are not networks. Access to or information about most of these data is already provided by GEONET. The binding constraints include the lack of efficient data base management systems for very large data bases and distributed data bases, inexpensive but capable work stations, and expert systems to simplify the use of complex data bases by scientists.

Dr. FISCHER. We have prepared a brief two-part demonstration lasting about 10 minutes.

The first provides examples of the use of computer technology to integrate and interpret multiple data bases, and the second illustrates one recent and important example of the uses of LANDSAT imagery.

Our expert in these subjects is Dave Nystrom, who is the Chief of our Office of Geographic and Cartographic Research.

Senator GORE. Very good. I guess we need to turn the lights down for you.

Mr. NYSTROM. We are going to try something a little different here. I am going to stand. I have to stand up to see the screen and watch Larry working the computer to make sure nothing goes wrong there.

Dr. Fischer mentioned our GEONET Telecommunications System across the country, and we have it on the computer, but to speed things up, I will just show you with a slide. I apologize.

Senator GORE. What is the capacity?

Mr. NYSTROM. The average links on that are 19.2.

Senator GORE. How many bits per second?

Mr. NYSTROM. 19.2 kilobits.

Dr. FISCHER. Slightly less than what you are recommending.

Mr. NYSTROM. We are using that system aggressively. We are communicating, and sending a lot of textual information. We have small project offices dialing in to do GIS work on the larger computers across the country. So it is working nicely. As we move into national and global initiatives like global change, clearly, when we start transferring the complex image data sets, that system is not going to be sufficient.

Senator GORE. It is not even close, is it?

Mr. NYSTROM. No. Right now when the EROS data center sends us, in Reston, an image data set, we send it overnight, and it might get there overnight on just one scene.

Senator GORE. What somebody called a graphic jam develops.

Mr. NYSTROM. And if it involves more than one scene, we use Federal Express.

Senator GORE. Well, that is all right.

Mr. NYSTROM. We are now going to jump into the fun part of this demonstration, which is the interactive part. And the first thing we want it to bring up are some projects across the country; and these are only projects out of our Reston USGS GIS research lab. We have research facilities in Denver; in Menlo Park, California; EROS Data Center, Anchorage; and other facilities across the country. These are only the projects out of our facility at headquarters.

The reason I wanted to point that out right now is to show the multi-disciplinary cooperative nature of what the technology is doing: a Maine transit project with Canada and the State of Maine; a new project with the National Capital Planning Commission right here, looking at resource allocation and architectural databases; a project with the EPA, looking at potential hazardous waste sites; the Bureau of Reclamation; Soil Conservation Service; the Fish and Wildlife Service up in Alaska; the Forest Service; the Bureau of Land Management—everybody—want to take their data and merge it with other agencies and do analysis and come up with results that we never dreamed of before.

Senator GORE. But you can do that now?

Mr. NYSTROM. We can do that now. Mike Nelson was nice enough to come out to Reston and help us choose examples . We have a wide range of applications for mining, forestry, water resources, geology, etc. We picked out two fairly short ones; one with the State of Connecticut, and then, if we have time, a transportation analysis out in California.

So we are going to zoom in now to the State of Connecticut. Larry has index maps set up for each one of these projects. Connecticut is a very typical project that we are dealing with. Connecticut was about ready to make a large investment in GIS hardware and software. We had done a lot of work with them looking at geology, hydrology, base category data, and we had other agency data sets in this part of Connecticut and, another important point, they had some real natural resource problems.

Every single one of those projects, none of them are hypothetical, is addressing real natural resource and environmental problems. We digitized around 30 data sets in this project area.

Just to give you a frame of reference on what that area looks like, Larry is going to bring up just three data sets: the land use that we digitized, the transportation net laid on top of that, the existing USGS transportation net, and the USGS hydrography layer.

This is a growing area. Hartford, Connecticut is down there, Springfield, Massachusetts up here. We did an industrial siting scenario and we looked at a lot of water resource applications, one of which we are going to show you today.

There is a town of Somers up here. In this part of Connecticut they have private water companies that supply well water to commercial areas and residential neighborhoods. In that two quadrangle area, the project area, here are the water companies, with the commercial area here. This one that we picked, the Somers Water Company, has wells and pipelines going to residential areas.

And we are going to zoom in and do some step-by-step analysis, looking for pieces of property that should be good for drilling a new well.

This is the outline. We told the GIS we did not want to look too far away. It drew an automatic buffer around that. Then we are going to start to window in on the many data sets and see what kind of properties we come up with.

Land use is critical in almost all of these GIS projects. We did not want to put a well on an existing residential or commercial piece of property, so we said okay, GIS, give us a new derivative map of non-built-up land. It does this very fast. And it shows wetlands and forest lands, and cuts out quite a bit of property.

We worked with the State of Connecticut with their data and they identified three polluted streams, so we put a 100-meter buffer around those three polluted streams. The land around that will be cut out. We worked with their State environmental people. They had, on a computer, a database that identified six pollution sources. We wanted to clearly stay away from those, so we put 500-meter buffers around those.

We have results after each one of these steps, but to speed up the process, we are just going to bring up a couple. But you can see where the buffers around those known pollution sources are, and where the cut was made on that stream right there.

We will skip a couple of steps. The next two steps are very critical from a water standpoint; we mapped out the surficial geology of this part of Connecticut, but we were mainly interested in the texture, which is in the database on the computer.

So we said in this part of Connecticut a good texture material is course grain—actually, stratified drift material. The area of that coarse grain material is in green. And then the final step we went through was to determine the saturated thickness of that material. We knew the bedrock geology elevation, we knew the water table elevation. We wanted a saturated thickness of greater than 40 feet, which is in the orange and pinks.

So the final result we came up with was some small pieces of property that should, from a water standpoint, be excellent for drilling a well. We made one final cut. We said okay, get rid of anything less than 10 acres. Again, it is almost an automatic calculation for the GIS system to do that.

And then many people said okay, fine, you came up with some results; how do I know exactly where those pieces of property are? You can bring back any one of the data sets. We will bring back just three: the transportation network, the hydrogragly, the geographic names—every name that is on a USGS map we now have in the computer to be used as reference data.

Larry is going to zoom in and you can basically see what street corner that piece of property is on. It is not a solve-all. They may go out and there might be endangered species there and you had not asked that question, or the person may not want to sell the property, whatever.

But that should be an excellent piece of property for drilling a new well.

One other thing we did on this project shows a different way to use the GIS, rather than the step-by-step process that we just went through, which is to actually interactively ask questions of all those 30 databases.

Another important point, Connecticut had been collecting data and putting it on computers for years. So they had the known pollution sources with all kinds of information; they had active wells with all kinds of information. But if you asked where are the active wells? Where are the known pollution sources? You would get printouts and the coordinates of where those were. But they never could spatially display the different departments' data sets.

Senator GORE. Well, I think this illustrates that, while the information may be available in columns of numbers or long books of words, you cannot possibly deal with that quantity of information unless you can present it in a visual form, so that our minds can cope with it. When you are presenting that volume of information in a visual form, that means you have to have the capacity to deal with lots of bits of data and transmit lots of bits of data from one computer to another; correct?

Mr. NYSTROM. That is correct.

Okay. What we are going to do very quickly is to zoom back to that Somers Water Company area. And we should see, if the last presentation was correct, six known pollution sources that we saw before, but there is an active well virtually surrounded by known pollution sources. So we, with the State of Connecticut, went in and said okay, tell us which well that is. All this is on a computer and had been for years, and they were never able to do this. It is the Preston Well, it is an active well. It is in stratified drift material. It is 44 feet deep. And years of flow data that they had collected on that one well, they can do that now any place in Connecticut.

We are going to go up and see what that known pollution source is. It is a former waste water lagoon, now inactive. But a lot of people say okay, that really looks close, but I do not know what my scale is any more. It is very, very easy to just put the cross-hairs on any two points and the system tells us, in this case, it is 209 meters away. And then, wherever the cross-hairs are, you can get the latitude, longitude or x-y coordinates of any features on the map.

Another concern they had was if that is a problem water. Well, let us just look at a couple of other data sets, like what is the land use around that well. It tells us it is a deciduous forested wetland. That gave them some concern because that goes up close to that former wastewater lagoon, so they said let us see if there is a correlation with the surficial material.

We called that up and it is a swamp deposit over sand, and clearly goes up where that former wastewater lagoon was. If that was a problem well and I worked for that water company, I might want to go take a look to see how that was cleaned up.

The State of Connecticut was so thrilled with the results of some of these projects that they instituted a couple new water resource processes after this project and they came back to the U.S. Geological Survey and said we are going to get GIS hardware and software; we need your data. And we signed a 50-50 co-op to digitize the state. Connecticut will be the first State with all base category data completed for transportation and hydrography at the scale of this pilot project. That should be done this year.

We can go to California on about a four-minute transportation presentation.

Senator GORE. All right. Let us look at that.

Mr. NYSTROM. The USGS has completed now the digitizing of the transportation net at one-to-100,000 scale and has turned that over to the Bureau of Census for the 1990 census. So just about every road has to be on there for the census-taking. Census is now adding to every road the name of the road, the addresses, so they can tie their census information to it.

This is a sample data set in Menlo Park, California—San Francisco being up here [indicating], the Bay up here [indicating]. We have a USGS Western Region office right here [indicating].

Say you wanted to get from A to B. The computer will show you the fastest way because Larry has gone in and given a speed limit to every class of road, from like 55 on the freeway to five miles an hour on, say, an alleyway. So it shows you the fastest way, but it also gives you the directions.

It says you started at Middlefield Road. You go 34 seconds, you turn left onto Willow, you go 141 seconds, you turn right onto

O'Brien. Once Census completes their task for the 1990 census, we can do that anyplace in the country.

Senator Gore. Domino's Pizza could use this.

Mr. Nystrom. A lot of people. And say it was an ambulance dispatch and somebody called in and said there is a terrible accident on one road and you still had to send the ambulance from A to B. You go in, you put a barrier on the road, you go back and ask the same question—what is the fastest way—knowing that there is a problem.

And it picked out in green another route going across the other interchange, telling you the time difference: And you can add all the other attribute information like one-way streets, traffic signals, and everything else to that data base.

Okay. We had better stop on that. There are a lot of applications of that to the U.S. Geological Survey in our hazard reduction work that we showed Mike when he was out at USGS.

And then the final four slides that we are going to show you are some LANDSAT slides of an application that the EROS data center worked on with the National Park Service looking at the fires in Yellowstone National Park. They are a little bit different size slides; we are going to have to zoom in and refocus here.

Now we are showing you this with slides because that is another technology problem at the current time. The interactive demonstration was a vector-based point line and polygon type GIS system; this is raster information which that system cannot handle.

Yellowstone boundaries are very faintly indicated in red. It is a LANDSAT image that shows the entire boundary of Yellowstone National Park, Yellowstone Lake, Old Faithful up in here. This is in July 1988 when the fires had just begun. Up in here you can see the smoke and in a couple of red areas down in here.

The National Park Service came to the EROS data center and said we have to get a handle on this. We have no good maps. So we went through a progression of LANDSAT images working with National the Park Service.

The next one is in September when the fire is raging and close to some very critical areas in Yellowstone, but clearly the best fire map that the National Park Service had at the time to get a handle on that.

The last one in time is in October 1988, when the fire had gone out, and you can clearly see the burned areas.

The final thing they asked for from the EROS data center was can you do a classification to pick up the most severely-burned areas so we can start our rehab management plans and feed that into a GIS system.

The last slide is such a classification done all digitally on an image processing system, the orange and reds being where the fire actually burned the crown cover of the trees.

So that is just one of many, many applications of GIS and imaging processing systems.

Senator Gore. Just a brief comment on this. Dr. Frank Press has made me aware that scanning through information of this kind in a systematic way makes it possible to forecast many volcanic eruptions and some earthquakes and thus avoid tremendous loss of life in many parts of the world. There have been some notable examples of where that could have been done if we had the system for looking at the information.

We are going to have a related hearing later in the year on that subject.

Thank you very much for your presentation, Mr. Nystrom and Dr. Fischer. Our final witness on this panel, Dr. Daniel Masys, is Director of the Lister Hill National Center for Biomedical Communications at the National Library of Medicine. Welcome and please summarize the high points, if you would, because in addition to our other time constraints they have informed us we will have a vote on the floor within about ten minutes.

STATEMENT OF DR. DANIEL S. MASYS, DIRECTOR, LISTER HILL NATIONAL CENTER FOR BIOMEDICAL COMMUNICATIONS, NATIONAL LIBRARY OF MEDICINE

Dr. MASYS. Thank you, Mr. Chairman.

On behalf of the National Library of Medicine and its Computer Research and Development Division, the Lister Hill National Center for Biomedical Communications, I am pleased to speak to you about current and future uses of computers and digital communications in the life sciences.

While my remarks will focus on the interests of the National Library of Medicine, many other components of the National Institutes of Health also play active roles in the development and use of computers in biomedical research.

[Slides shown.]

Dr. MASYS. The NLM, the National Library of Medicine, has a long and productive history of innovation in the use of computers in biomedical research. It was a quarter of a century ago, in 1964, that the NLM first created a computerized version of the Index Medicus, an index catalog to the world's biomedical literature.

Seven years later, in 1971, this data became available for on-line searching over public computer networks in a system called MEDLINE. Over the past two decades, the NLM has made its computer data bases ever more widely available over commercial value-added networks.

Now, access to the Library's computer systems is provided via five different U.S. telecommunications vendors. Today MEDLINE and the 38 other data bases that comprise the system called MEDLARS are the largest and most widely-used biomedical computing system in the world.

Over four million computer searches were done on the NLM computer system this past year. We know that about half of those searches were done for medical research and education, and the other half were to get information for the direct care of a sick patient. By all measures, the MEDLARS system has been a spectacular success, the only practical way that a biomedical researcher or health care practitioner can keep up with the vast and growing knowledge in medicine and biology. Over 300,000 new articles are added to the system each year.

In addition to these public services over commercial networks, the Library has a vigorous R&D activity. The Lister Hill Center and the recently-created National Center for Biotechnology Information are both connected to the National Research Internet via a MILNET node at the Lister Hill Center and to a Cray supercomputer facility run by the National Cancer Institute.

Let me review, though, for just a moment the type of information that is traveling over those networks. MEDLINE, as you have heard in other cases this morning, is a bibliographic data base. That is, each record in the system is not the original publication but a synopsis of the original—containing the title, authors, key words and such.

Each record averages about 400 words, about 2,000 characters, and this record size matches well with the capability of the current generation of commercial computer networks, which can generally carry up to about 240 characters per second over standard telephone lines. Thus, even as we are speaking today thousands of literature citations are being transmitted anywhere in the U.S. to individual doctors' offices in, say, Tennessee or Virginia, within just a few seconds.

But biology and medicine depend on much more than words. Much of the understanding of complicated processes of health and disease lies in images, pictures of body systems and organs and molecules which cannot effectively be described in words.

A standard black and white x-ray such as this skull film has the equivalent of at least ten million points of light or pixels. We can and do make electronic versions of such images at the NLM, but to transmit these images over the currently-available computer networks is simply not feasible. It would take more than ten hours to send just one x-ray picture.

There are currently two areas of biomedical research which have a compelling need for high speed computing and computer networks. They are biotechnology and computer-based medical imaging.

As you know, DNA is the central focus of biotechnology. This long molecular thread, present in the nucleus of each of our ten trillion cells, guides the assembly of our bodies and controls the maintenance of all life processes. Molecular genetics scientists have determined the exact sequence of the DNA of nearly a thousand human genes, yet this represents less than one percent of man's genetic information.

Like the rapid escalation of aerospace technology that led from Kitty Hawk to the moon in less than 70 years, however, the pace of acquisition of this vital information about inner space is accelerating dramatically, and in order for society to benefit from the increasing rate and volume of this molecular biology information great advances will be needed both in methods of computing and in digital communications.

Senator GORE. It is doubling every eight months, last I checked. Is that right?

Dr. MASYS. The doubling time is continuing to shorten. It is now slightly more than six months.

As an example, GENBANK is the National DNA research data base. This is a GENBANK record, a very complicated one, for a cancer-causing gene, a so-called oncogene. The actual DNA sequence is represented at the bottom of this screen, the actual As, Cs, Ts, and Gs of the DNA sequence.

The critical life function of molecules as complicated as this, however, is not dependently solely on that sequence but rather on the three-dimensional structure of the molecules that are coded for by that DNA.

There exist models for predicting how those molecules fold, but when these models are currently applied to molecules with hundreds or even thousands of atoms the problem becomes trans-computable. It is simply too difficult to do, even with some supercomputers, where it requires years of supercomputer time.

Massively parallel architectures are one hope of solutions of these very complicated molecular problems, and if we are able to solve them it provides the promise of dramatic new ways by which we can both analyze disease-causing agents, such as the AIDS virus or cancer-causing agents, and the therapies that will interact with them.

At the same time that this molecular science is advancing rapidly, new technologies have appeared in clinical medicine to make detailed pictures of our inner body composition.

Computerized tomography is one now well established in medical practice and is being supplemented by new things, such as magnetic resonance imaging and positron scanning. As you know, each of these technologies yields a series of two-dimensional pictures as if we had slices through the internal anatomy of our body.

It is now becoming possible with high-speed computing, however, to reconstruct the original three-dimensional internal anatomy that these pictures represent. These three-dimensional images can be rotated, examined from all angles, and can even be used to create replacement body parts, such as bone implants. Using research tools, surgeons can get a clear mental image of what they will discover at the operating table and can even experiment with various surgical approaches on electronic images.

I would like to show you now a brief video that has an example of three-dimensional computer reconstruction. This is work under way on an NLM grant at the University of Washington in Seattle; Dr. Cornelius Rosse is the narrator.

[Videotape shown.]

Dr. **Masys**. That is the end of our video.

To compute the complex three-dimensional images you have just seen from mathematical data representing natural structures in the body and moving them in real time requires truly massive computing power and specialized computer architectures.

For the foreseeable future, such specialized computers will be available only at a relatively small number of sites, and access to these machines will require high band width computer networks capable of moving billions of bits per second, representing those complex medical images.

In conclusion, it is clear that in basic research to understand both the molecules that make us human beings, that control health and disease, and in the exploration of living human anatomy in ever finer detail, the life sciences research community will be an important beneficiary of progress in computer science.

There is no doubt that high speed computing and digital communications are an essential technology which will pay dividends in both accelerating biomedical discovery and improving the health care of our Nation.

Thank you.

[The statement and questions and answers follow:]

Testimony by

Daniel R. Masys, M.D.
Director, Lister Hill National Center for Biomedical Communications
National Library of Medicine

National Institutes of Health
Public Health Service
Department of Health and Human Services

On behalf of the National Library of Medicine and its computer research and development division, the Lister Hill National Center for Biomedical Communications, I am pleased to speak to you about the current and future uses of computers and digital communications in the life sciences. While my remarks today focus on the interests of the National Library of Medicine, many components of the National Institutes of Health also play an active role in the development and application of the most advanced concepts in computing to biomedical sciences. New computer and communication technologies will have a wide ranging impact on research presently supported by the entire National Institutes of Health.

The National Library of Medicine (NLM) has a long and productive history of innovation in the use of computers and computer networks. It was twenty five years ago, in 1964, that the NLM first created a computerized version of Index Medicus, an index catalog to the world's biomedical literature. Seven years later, in 1971, the NLM established a computer network communications contract with a fledgling company called Tymnet, to provide long distance digital telecommunications for online literature searching of that Index, via a new system called MEDLINE. Over the past two decades, the NLM has made its computer databases every more widely available over commercial value-added computer networks; access to the Library's computer systems is now provided via five different U.S. telecommunications networks.

Today, MEDLINE and the more than twenty other databases which make up a system called MEDLARS (Medical Literature Analysis and Retrieval System), comprise the largest and most widely used biomedical computer system in the world; over four million computer searches were done on the NLM computer system this past year. We know that about half of those searches were done for medical research and education, and the other half were to get information for direct care of a sick patient.

To simplify access to MEDLINE, the NLM makes available a microcomputer program called Grateful Med, which allows a health professional or researcher to compose a database search locally, by simply filling in a form displayed on the pc screen. The program then automatically connects to the NLM computers across the value added networks, conducts the search and downloads the results to the user's own computer. Over 10,000 copies of this program have been distributed, and the majority of searches of NLM databases now are conducted via this "user-friendly" software. Grateful Med provides users access not only to MEDLINE, but also to a growing number of specialized information collections on topics such as AIDS, cancer treatment, bioethics, and toxicology databanks. Extensive data on the medical and environmental effects of hazardous chemicals is available via the TOXNET collection of databases, including recommendations on management of emergency spills and other environmental releases.

By all measures, the MEDLARS system has been a spectacular success, the only practical way that a biomedical researcher or health care practitioner can keep up with the vast literature in the life sciences. Over 300,000 new articles are entered into the system each year. To put this into perspective, if your doctor were conscientious and took the time to read two new articles in medical journals each evening, at the end of a year he would have an 800 year backlog of material to read. Systems for searching the literature by computer allow him to find the latest and most useful research findings within seconds, among the millions of publications in the system.

In addition to its public services over commercial networks, the Library has a vigorous research and development activity. The Lister Hill National Center for Biomedical Communications is the computer R&D division of the Library, and the recently created National Center for Biotechnology Information is developing new computer methods for managing the data of molecular biology. Both research groups are connected to the national research Internet via a MILNET node at the Lister Hill Center, and to the Cray supercomputer facility run by the National Cancer Institute. Connections between computer networks, known as "gateways" have also been established between the NLM and the American Medical Association's AMAnet, to provide AMA member physicians instantaneous access to the Library's databases.

Let us review for a moment the kind of information which is travelling over these digital networks. MEDLINE is a bibliographic database. That is, each record in the system is not the original publication, but a brief reference to the original, containing the title, authors, source, and subject listings for each publication, often with an abstract of the article. Each record averages about 2000 characters, and this record size matches well the capability of the current generation of commercial computer networks, which carry up to 240 characters per second over standard telephone lines. Thus, a complete literature citation can be transmitted nearly anywhere in the U.S. within a few seconds.

MEDLINE contains words and numbers describing the literature of the life sciences. But biology and medicine depend upon more than words. Much of the understanding of complicated processes of health and disease lies in images, pictures of body systems, organs and molecules which cannot effectively be described in words. A standard black and white chest x-ray contains the equivalent of at least 32 million points of light, or pixels. We can and do make electronic versions of such medical images at the NLM, but to transmit these images over currently available computer networks is simply not feasible. It would take more than ten hours to send just one x-ray picture.

There are currently two areas of biomedical research which have a compelling need for high speed computers and computer networks. They are biotechnology and computer-based medical imaging.

Biotechnology

As you know, the genetic code contained in the nucleus of each of our cells is Nature's language of life: the mute but eloquent set of instructions which guides the assembly of our bodies, and controls the maintenance of all life processes. A long molecular thread called DNA is the chemical parent of us all, and over the past twenty years we have taken our first halting steps towards comprehension of the messages encoded in the more than three billion units which comprise the human genome. Molecular genetics scientists have, through laborious manual methods, determined the exact chemical structure of nearly 1000 human genes, yet this represents less than one percent of man's genetic information. Like the rapid escalation of aerospace technology, however, which led from Kitty Hawk to the moon in less than 70 years, the pace of acquisition of this vital information about "inner space" is accelerating dramatically. In order for society to benefit from the increasing rate and volume of new knowledge in this field of Biotechnology, great advances will also be required in methods of computer storage, retrieval, analysis, and communication.

As an example, there is a database built at the Brookhaven National Laboratory which contains the structures of proteins. It has been possible since 1965 to turn the mathematical values in the data base into stylized pictures, at least for small molecules. With current day computers, larger molecules can be modeled, using color 3-dimensional displays. Such analysis often gives clues as to critical parts of the molecule, where it binds to other molecules and tissues within the body. Two-dimensional NMR techniques are now leading to analogous structures in solution and these techniques promise to expand the database rapidly.

GenBank is another famous research database, which contains the sequence of the DNA which codes for proteins. We now have automated methods for determining the sequence of DNA and proteins, which may be thought of as pearls on a string. The critical life functions of molecules, however, result from the three dimensional folding and curling of the pearls on the string. There exist models for predicting how molecules fold, but when these models are applied to hundreds or even thousands of molecules such as are found in large proteins, the problem becomes "trans-computable"--requiring years of supercomputer time with today's programming techniques and computer architectures. Supercomputers and massively parallel computers provide one hope of overcoming this analytical roadblock, so that we may take a primary molecular sequence, and predict the three dimensional structure of the resulting product. But knowing the principles of efficient software design for such machines is, like the machines themselves, in its infancy.

If we were to succeed in better hardware and programming methods, how would this be useful? As an example, in the absence of a more efficient process, the Government and industry currently must, in many circumstances, randomly screen chemicals and compounds for activity against diseases. AIDS and cancer are two examples where literally tens of thousands of compounds have been randomly screened for activity against the disease. In contrast, AZT and some other agents were selected for screening because of structural characteristics. A generally useful method for molecular modelling would greatly improve selection and development of new drugs. If we could via computer-based molecular modelling predict both the structure of the disease causing agent, for example the AIDS virus, and the structure of the drug which would bind to and inactivate the virus, the development of vaccines and antidotes would be dramatically accelerated. Research is under way to improve the speed and accuracy of molecular structure calculations, and it is clear that this scientific domain will require the fastest and most capable of new computer designs.

Better methods of storing and analyzing molecular data by computer are under development at the Library's National Center for Biotechnology Information. In particular, a system called GenInfo allows a researcher to enter a question in his own words, as if speaking to a colleague. The GenInfo system takes that question and can search up to thirteen different molecular databases for the answer, displaying not only text and numbers, but graphs, chromosome maps, and full color pictures of molecules. The system is currently under testing at the National Institutes of Health, but wider access will require high bandwidth network connections -- the ability to move large volumes of information over long distances at high speed.

Improving communication facilities for exchange of experimental data and for access to central data bases is of great importance to many university-based research teams that are supported by the new National Center for Human Genome Research, the National Institute of General Medical Sciences (NIGMS), the National Library of Medicine and other NIH institutes. Much basic science university research work that is supported by NIH, such as the crystallographic study of large biological molecules, is highly dependent upon shared access to NSF supported supercomputers. Such access requires high bandwidth and would benefit in the future from enhancements in transmission capabilities.

Computer-based imaging

At the same time that molecular science is advancing rapidly, new technologies have been developed to make detailed pictures of our internal organs and systems. Computerized tomography is now well established in medical practice, and is being supplemented by new technologies such as magnetic resonance imaging and positron emission scanning. Each of these methods yields a series of two dimensional pictorial "slices" representing internal anatomy. It is now becoming possible, however, to take large numbers of two dimensional pictures and reconstruct the three dimensional structure which they represent. The three dimensional image can be rotated, examined from all angles, and can even be used to create replacement body parts, such as bone implants. Surgeons can get a clear mental image of the anatomy they will encounter at surgery, and can even experiment with various surgical approaches on the electronic image.

To compute complex three dimensional images from two dimensional pictures, and display them moving in "real time" requires a massive amount of computer power and specialized computer architectures. For the foreseeable future, such specialized computers will be located at a relatively small number of sites, and access to such machines will require the capability of moving millions of bits per second, representing complex medical images, over digital networks.

It is clear that in basic research to understand the molecules that control health and disease, and in the exploration of living human anatomy in ever finer detail, the life sciences research community will be an important beneficiary of better networking and more powerful computing methods. We at the Library believe it is important that the technical advances proven in high speed research networks make their way into commercial value-added networks, so that the NLM's public database services can be enhanced in the future to provide graphical and pictorial information. High speed digital communications is an essential technology which will pay dividends in accelerating biomedical discovery and in improving health care for our nation.

Answers to questions from Senate Committee on Commerce, Science, and Transportation

D. Masys, M.D.
National Library of Medicine
9/22/89

1. What kind of R&D is needed to improve the way we manage scientific data?

   Over the past two decades we have proven that we can dependably build and use specialized scientific databases. Each of these databases has generally had its own unique structure, vocabulary, and rules for searching. The questions of science, however, often range over a spectrum of different kinds of scientific data, so the challenge to computer systems designers and builders today is to create linkages among scientific data sources. These linkages are needed to allow related data from many different sources to be retrieved and analyzed in an efficient way. Much work is needed to devise and prove the feasibility of such "distributed databases." In addition, since data collections are growing to massive proportions, higher speed network connections are needed for the rapid exchange of data among researchers.

2. Which federal agencies are most involved in improving data management?

   The membership of the Federal Coordinating Council for Science, Engineering, and Technology (FCCSET) includes representatives from virtually all agencies within the government which sponsor or perform research and development activities. Information is the life blood of science, and also of government itself, so that within each federal agency there is a nucleus of staff whose focus is the improvement of data management. For the national scientific enterprise, major programs for data management have been undertaken by the Defense Department, NASA, the National Institutes of Health, the National Science Foundation, the Department of Energy, the Library of Congress, and the Department of Agriculture.

3. Has the Administration recognized the need for more funding in this area?

   The recent report on high-speed computer networking and advanced computing research released by the Office of Science and Technology Policy highlights the important role that computer and communications technologies play in fostering scientific achievement and industrial competitiveness for our country. Specific funding information is not available.

User-friendly software

You have discussed how Grateful Med was created to make MEDLINE easier to use. Yet, there are many databases that are so complex they can only be used by "information specialists" who look up things for customers.

1. Are the online systems evolving into user-oriented systems?

   Yes. This is a major movement throughout the database "industry", driven primarily by the advent of increasingly powerful personal computers which can perform many of the tasks previously performed by information specialists.

2. If yes, what is needed to accelerate this evolution (e.g., consistent structured user interfaces, the development of expert systems based on artificial intelligence, user interfaces incorporating graphics, better search algorithms)?

   All of the technologies mentioned will undoubtedly contribute to better information access by end-users. However, it appears that the greatest benefits will come from work done "behind the scenes" in the adoption of standards. This includes standards for the vocabularies used to describe the scientific information contained in databases, standards for command languages which work predictably and reliably in different computer systems, and standards for data interchange among different computer databases. Technical standards are, like human language, our agreement upon the terms by which we will communicate with one another. The end-user will benefit dramatically when he does not have to learn a new "language" or set of rules for every database that he wishes to access, and when computers can easily exchange chunks of data over networks using widely-accepted standards.

   In this regard, the National Library of Medicine is currently pursuing a research effort in collaboration with several university groups to create a Unified Medical Language System. This is not a new dictionary of medicine, but rather a translation system, an electronic Rosetta stone which can detect and use biomedical concepts as they are represented in a number of national medical databanks and as they are used by health care practitioners and researchers.

Networks versus optical disks

The Library of Medicine uses both networks and optical disks to distribute information.

1. What are the benefits of centrally-managed on-line databases, vis-a-vis optical disks.

   Until recently, only large mainframe computers provided sufficient data storage capacity to handle large databases; this is changing now with the advent of inexpensive optical media capable of storing hundreds of millions of characters. The principal advantages of central data resources are the efficiency of making available rapidly changing information, providing access to data collections which are larger than even the most capacious of optical disks, and providing analytical computing power greater than that locally available to users.

2. What are the down-sides of the two approaches?

   The time required to produce and distribute copies of data on optical disk makes it less attractive for rapidly changing applications (e.g. research results in cancer treatment, or newly-sequenced DNA molecules). The problem of ensuring that users have the latest and most complete version of a database is compounded when users have optical disks containing outdated versions of databases. Central systems, on the other hand, suffer from communications bandwidth limitations: it is generally far more expensive to transmit large volumes of data over communications lines than to retrieve that same volume of data from a local optical disk.

3. What is the appropriate technology mix, and what applications are appropriate for optical disk technology?

   Optical disk technology continues to be the high-capacity digital storage medium of choice for very large data sets. The best of online and local optical disk systems will be a hybrid of the two, embodied in transaction-oriented systems. This general model presupposes that many users will have optical disk systems attached to their local computer (e.g. a personal computer with a CD-ROM player). When they initiate a database search, the computer will first search there local optical disk database, and then if needed, go online to a central databank. The search of that online databank will in effect be a statement that says "send me only the information which has changed since my optical disk was released." Such system designs make the best use of low-cost, high capacity local optical disks, and the currency and completeness offered by central databanks, while minimizing the amount of data which must be transmitted over communications lines.

Coordinating Federal data management

1. What can be done to improve coordination of federal computer systems and communications networks? What is most important?

   As in the private sector, the best overall utility for users of computer systems and networks results when systems are developed using widely-accepted, vendor-independent standards. Promoting the use of interoperability and communications standards where appropriate within the government would provide an important basis for improved coordination.

2. What are the principal issues in ensuring that users can effectively and conveniently access and integrate information from a variety of data bases and a variety of networks?

Just as there was no national railway system until agreement was reached on the gauge of the tracks, agreement is needed on the protocols by which computers communicate with one another, how they "package" the data for transmission to other computers, and how they interpret the commands which retrieve appropriate information. The user interfaces of computers need to provide intellectual assistance like that provided by professional search intermediaries to assist users to formulate questions in their own words; programs will then translate those questions into the appropriate vocabulary and syntax for extracting data from databases, whether those databases are at the users own machine or are at a distance, accessible over long distance networks. We are very much in the era of stand-alone databases in this regard, and much technical and organizational work remains to create a seamless fabric of national information sources.

Senator **Gore**. Thank you very much. That was a fascinating presentation.

I can say from personal experience that when a child is injured and the doctors are trying to decide what surgery might be needed, you are glad that they have the ability to see a three-dimensional presentation of what they are going to encounter when they get inside during the surgery. That just dramatically changes the range of options available and has tremendous benefits for the patients.

The experience I was alluding to, a member of my family faced surgery and they were able to use imaging to examine the thing that goes around the kidneys?

Dr. **Masys**. The renal capsule.

Senator **Gore**. They had another name, too. It was longer and more difficult to pronounce. But in any event, they were able to determine that they did not have to do surgery in that particular area, and that is a good thing, because experience shows that once they go in there, it is Pandora's Box and very difficult to ever get back to normal again.

So I have a personal appreciation for this.

Incidentally, later this year, our subcommittee is going to have a hearing on the human genome project. The amount of data, which you say is doubling in just over six months, now?—the implications of the work is just staggering.

I had a personal experience, I want to note, as well. When I got out of the Army in 1971 and was in graduate school, I was doing some research on a biomedical ethics issue at the Vanderbilt Medical Library and I was absolutely amazed at the difference between the way information was presented there and the way it was available in other libraries.

If you wanted to do a research project, utilizing information in "The New York Times" or "The Washington Post" or the periodical index, it was a very difficult process involving lengthy delays between the publication of information and its availability in the library.

But what I found 18 years ago, as a result of your institution's work, was that new publications in medical journals are almost instantly available, at least in abstract form, through this computerized system. I was absolutely amazed. You all were into this long before most other institutions, and I just want to add my compliments there.

We face a time problem here. I am going to recognize Senator Robb. I do that with apologies, but we have a third panel, we have a vote on the floor, and the day is getting old.

So what I am going to recommend is that I will put my questions in the record for you and when we come back after this vote we will move on to the third panel.

At this point, we do have a few minutes here, and I want to recognize Senator Robb.

Senator **Robb**. Mr. Chairman, I would be happy to defer, too. Let me just ask one general question, again to increase the layman's understanding.

I assume that all of the input that is necessary to make these computer models and computer graphics, is still the most time-consuming part of the process? I mean, once you get the information in, the computer does the rest of the work.

Is it safe to assume that it is more valuable for applications where the input data are essentially static and do not change a lot? I was thinking of the comparison between the environmental model that was shown first, and comparing that to the traffic model, where you would have to put in things like accidents, rush-hour—all of the variables that would ultimately affect the answer at that particular time.

In other words, if there are lot of important variables does using this type of information becomes less cost-effective. Is that a fair assumption? I am trying to understand the application to other situations.

Mr. NYSTROM. I will be happy to address that from a geographic information standpoint, since you use those two examples.

We have already heard about standardization, and classification is a part of that. If, say, two towns out in California added information to a transportation net but they did not coordinate, and you fed those two databases into a GIS to do a kind of regional analysis, the GIS system will not [even] know how to deal with the different classifications. So standardization of data is critical on all types of natural resource, environmental, medical, geological applications.

We have a long way to go because the National Park Service, the Forest Service and the Bureau of Land Management, for example, are going to have to start calling forestry classification things the same to do an analysis on that whole Yellowstone area along with the State. If they do not use the same classification, a GIS system cannot understand those different databases.

Dr. MASYS. I would just say briefly, it is clear networks have their highest value when information changes fast or is of very large volume, and in the medical domain it is interesting, however, that we have two bottlenecks. We have these clinical scanners driven by computers, which generate massive data sets, which then have two bottlenecks.

One is, to do the kind of rendering you saw on that TV screen takes hours with current computers, and the other thing is to try and transmit those images across the country, where there may be a specialist who knows that particular rare disorder.

Similarly, it is faster to use a reel of tape and a jet airplane right now, than to move that over a network. So we have a number of challenges in medical data.

Senator ROBB. Mr Chairman, I recognize the time constraints. I see there is a vote on.

Senator GORE. Just to follow up from what you said there, there is a joke in the Soviet Union—they have a tremendous sense of humor about their predicaments from time to time—about a decision to change what side of the highway they drove on. The upshot of the story is that they decided to compromise and have the trucks do it the first year and automobiles the next year.

Somebody has used the phrase, "a graphic jam." What we have here when different databases and different standards and different formats, it is almost like the cars running into the trucks. It makes the entire information highway system unusable. That is why we need a national approach and a national network, and a continuing effort to address these problems of standards and access and so forth, when they come up.

One final point. I used an example, Senator Robb, right at the very beginning of the hearing, that builds on this LANDSAT database.

It was amazing to me to find out that after all the money we have spent taking these pictures and getting this information stored, 95 percent of the pictures taken of the earth's surface by LANDSAT have never been seen by a single human being. Yet if we can use the kind of data management that was demonstrated here, we can very quickly pull out useful information from virtually all of the stored pictures.

Well, we had a lot of detailed questions about some of the technical issues and some of the policy issues. Again, apologies for not spending more time with you directly in questions and answers, but when we come back after this vote we will go to panel 3.

Again, to the members of this panel, thank you very much. We appreciate it. We will be back in about in about ten minutes.

[Recess.]

Senator GORE. Welcome. Let me welcome our third panel, Dr. John Seely Brown, Vice President for Advanced Research at the Xerox Palo Alto Research Center; Mr. Ted Nelson with Autodesk, Incorporated in Sausalito; Dr. Irving Wladawsky-Berger, Vice President and Site General Manager with IBM in Kingston, New York; and Mr. Richard T. Wood, Senior Vice President for Business Development of University Microfilms, Incorporated, in Ann Arbor.

Let me announce that during the break, Senator Bryan and Senator Robb were added as co-sponsors of S. 1067, so we made a little progress on the bill even during the break. I am pleased and grateful to them for joining as co-sponsors.

I know that members of this panel have been made aware of the limitations on time. I apologize for the fact that we got so absorbed in the presentations of our earlier panels that we used a good bit of time, but I want to assure you that we are greatly interested in what you have to say.

If you can summarize it, we will be looking carefully at the full texts of your prepared statements, but at this point, if you can make such presentation as you feel is appropriate to get at the essence of what you have to tell us, we would appreciate it.

We will begin with Dr. Brown. Thank you very much for being here. Please proceed.

STATEMENT OF JOHN SEELY BROWN, VICE PRESIDENT, ADVANCED RESEARCH, XEROX PALO ALTO RESEARCH CENTER

Dr. BROWN. Thank you very much, Mr. Chairman. It is a great opportunity to be here.

Very briefly, I would like to focus on what I consider to be the major theme for our attention today and that is: how can we capitalize on the Nation's knowledge better, knowledge both in the forms of the explicit knowledge that is recorded in the archives, in our books in our libraries and so on, and also the knowledge that is tacitly held presently in our own heads.

More concretely, how can we make better use of what we know? This is the major challenge, and I see it as a major reason for the informational infrastructure you are really pushing for.

What I would like to do is focus my comments on two issues. First, how can we actually come to understand the role of documents in our society and use them more effectively. Second, a topic that has not come up thus far today is, how can we use such informational infrastructures to radically facilitate more effective collaborative work?

First, think about documents. Why should we find documents interesting in the first place? Why should we care about how to store, retrieve, or manipulate information in the form of documents?

I hold there are two reasons. The first is to recognize of course, that documents are really the medium in which our culture makes advances and in which we are able to record our accumulated knowledge.

Similarly, documents can be viewed as the glue enabling organizations, corporations, to function. In some ways, documents are almost like an interchange protocol people use in working more effectively together.

So documents are very important. And documents are going to become even more important as we get the right types of informational infrastructures in place.

But the curious thing is that from a technological perspective, we tend to confuse documents as just being simple ASCII text. This morning you saw dramatic examples to the contrary, (e.g., dynamic images and animation), but in fact even the most popular types of documents—books and papers—are really quite different from simple ASCII files.

Documents include objects, graphics, images, text, layout, font, et cetera, et cetera, structured in order to facilitate human comprehension--human comprehension! I want to concentrate on this.

Curiously, you all know the adage: a picture is worth 1,000 words. This is an absolute truism from the point of view of how the human mind functions. However, from a technological perspective, it is just the opposite—basically a picture image of those 1,000 words of text takes 50 to 100 times more in storage space for a computer system than 1,000 written words of text would consume. So what is optimal for the human mind turns out to be not very optimal for current technological devices in terms of storage and band width requirements.

The informational infrastructure you are proposing now enables us, perhaps for the first time, to give justice to documents of old and documents of new.

I say documents of new, because it is clear that documents today are becoming much more image-intensive. For example, today we saw various types of multimedia documents, video documents were shown here; and the demands those types of documents put on our informational infrastructure are even greater.

In order to bring a marriage between the way humans can best understand information and how to find the information they want, we have to rethink the capacity requirements for informational infrastructures. Your bill does this.

The second point I wanted to present is the notion of collaborative work. Now, why do we care about collaborative work? In some sense, collaboration is probably the way we achieve most of our creativity. Also, collaboration is how we in fact conduct most of our daily affairs. Ironically, I think most of us in this room spend at least 50 percent of our time in meetings and perhaps the other 50 percent, like myself, in travelling to meetings, using classical types of infrastructures—roads, planes, etc. Somehow, we have to find a way to break that bottleneck.

Senator GORE. What bottleneck is that, again?

Dr. BROWN. The bottleneck in having to come together in order to have effective meetings and effective collaboration.

Senator GORE. Come together physically in order to come together mentally?

Dr. BROWN. Right. So what we really need to do is think about a use of your informational infrastructures in order to achieve a kind of co-presence without actual presence. I think if we could do that, our productivity could go up dramatically.

Senator GORE. That is where they come up with this word collaboratory.

Dr. BROWN. Correct. I think we have focused on collaboratory discussions in the past primarily from the scientific or research community point of view.

I want to dwell on two different aspects of that notion of collaboration or collaboratory. The first comes from the perspective of a large corporation. A multi-national corporation like Xerox tends to have design teams of 50, 100 or 150 people working on projects with people spread all over the Nation, or all over the world. It is an incredibly inefficient operation to try to bring coordination using current technology to make teams optimally synergistic.

The kinds of informational infrastructures we are talking about here could radicalize the way we could make decentralized design teams, product teams perform effectively.

Senator GORE. So you are giving us a real world example of how your company, Xerox, could better compete with others around the world, by being able to have your research and design people in various physical locations communicate with each other with high——

Dr. BROWN. With high band-width information channels.

Senator GORE [continuing]. Volumes of information going back and forth so that they could collaborate without everybody getting on the airlines and flying all over the place and trying to coordinate their schedules and so forth. In that way, they would be able to mentally work together and collaborate without everybody flying to one physical location.

You think that would help your ability to compete?

Dr. BROWN. Absolutely. I think it would help our ability to compete and the Nation's ability to compete. I think we waste a tremendous amount of time.

Senator GORE. You are pretty certain of that?

Dr. BROWN. I am pretty certain of that fact.

A good share of our research in terms of how to streamline our own internal processes are aimed now at finding radical ways to bring about effective collaboration, which turn again on the effective use of the band-width that your type of national infrastructure could provide.

In summary, let me make a pun.

Senator GORE. I know the answer to this, I think, but for the record, why can your company not spend the money to put in place a three or four or five gigabit connection between your various locations?

Dr. BROWN. We can spend the money to do some very simple prototypes that use fairly low band-width nets, but, basically, we cannot leverage the kind of vast infrastructure you are talking about operating at all our different branches all over the country and all over the world and also interact with our supplier corporations.

Senator GORE. Just as a delivery company could not build the interstate highway system.

Dr. BROWN. Not and stay in business. Right.

In summary, I think it is interesting to step back and recognize we are really looking at two types of collaboration here. The first part of my statement deals with how do we more effectively collaborate with the past; how do we rapidly interact with the archives; how do we rapidly access what is already known. That is a collaboration with the past.

On the other hand, how do we more effectively collaborate with the present in terms of being able to achieve a sort of co-presence without physical presence?

Thank you.

[The statement follows:]

Statement
by

John Seely Brown,
Vice President, Advanced Research,
XEROX Palo Alto Research Center (PARC)

CAPITALIZING ON THE NATION'S KNOWLEDGE: MAKING USE OF WHAT WE KNOW

INTRODUCTION

In its potential to amplify productive and innovative practices in the crucial years of international competition that lie ahead, the National Research and Education Network offers an extremely exciting prospect. Along with its generally recognized possibilities for research and education, we find latent within this high-performance network means to overcome some of the present and future challenges to the nation's entire manufacturing and technological base.

The immense power of information technology to help us to meet the nation's demand for knowledge is unquestionable. But the *information explosion, increasing complexity*, and the *ever accelerating pace of change* are already overburdening "knowledge workers." From our perspective, we see a high-performance network, augmented with the appropriate knowledge tools, as both helping to meet these general challenges, and also providing the nation's research and manufacturing base with the means to recognize, value, and capitalize upon the nation's knowledge.

Two sorts of work practices that are crucial in all areas of society in the age of information technology can be successfully supported by a high-performance network. One is *collaboration* --- the synergistic work practices of groups of people. And the second is *interpretation* --- the synthesis of raw data into something intelligible. The testimony that follows attempts briefly to highlight the significance of these practices and to give some indication of how an effective network can support them.

Statement by John Seely Brown

SUPPORTING COLLABORATION AND INTERPRETATION

1. Leveraging the nation's knowledge base --- Empowering collaboration

If we have been able to see further, it is by standing
on the shoulders of our predecessors

A well designed National Research and Education Network will be as important and significant for advancing the nation's scientific, technical, manufacturing and cultural resources at the end of the century as the advent of the public library system at the beginning and the development of the national highway system in the middle. Libraries are not, it should be remembered, merely passive repositories of information, they are active centers with formal and informal resources to help people gather, develop, and exchange knowledge. Similarly, the highways are not merely a means of travel, but also an important step in allowing separate places to interact productively. The promise of a sophisticated network, then, lies in its potential to be not just a means for exchange and retrieval, but also the locus of synergistic collaboration.

Collaborative groups are particularly powerful sources of insight and learning. They provide resources that are more than the sum of their parts. The Japanese, in particular, have shown us the value of collaborative work. To meet the demands of the future, we need not only to amplify collaborative practices in the local workplace, but also to enable them to occur in networks across the nation.

Until now, space and time have been effective barriers to good collaboration. But if we can understand how to bridge them efficiently with information technology, we can build a national collaborative database and infrastructure --- a Collaboratory --- that can bring people and resources together across time and space to allow them to work productively in ways that have not previously been possible.

In particular, a national network augmented with knowledge servers can unite individual talent with organizational resources. The MOSIS project was one instance of this. Student chip designers were given access to the knowledge servers that assembled designs into packages that could be cost-effectively produced in VLSI process lines. This is only one example. Many more could be given to show how such direct access could rekindle the passion for design (mechanical, chemical, electrical, biotechnological, etc.) among students and designers in our society. Moreover, this system could rapidly spread current approaches to design through the technological infrastructure and thereby keep industrial and other designers, the half-life of whose skills is continually falling, in active touch with contemporary knowledge, practices, and developments. Thus the network could be used to update the skill base of technical workforce, which now often has an active life of only five or fewer years. Such accessible learning resources are becoming increasingly vital to the survival of our rapidly changing manufacturing base.

A high-performance network will give large organizations access to the creativity and independence of small suppliers and individuals outside the organization and in other parts of the country. It will thereby help build new and valuable webs connecting large and small in important strategic alliances. The concentration of creative talent pools in major metropolitan areas will then not be such a problem for the development of other areas. In connecting the large and the small, we may be able both to unleash and to distribute more evenly the nation's creative resources.

Two vectors of collaboration --- standing on shoulders, not toes

Productive, synergistic collaboration is already a key, though often overlooked, ingredient of effective working. It has two vectors. In one obvious direction, we interact with contemporaries --- communicating with colleagues, forming alliances, and thereby being elevated onto their shoulders, not standing on their toes. In the other less-recognized direction, we collaborate actively with the past, drawing on the wisdom of predecessors to provide a foundation for our ideas. In constitutional theory and law, for instance, understanding historical precedent is immensely important. It is less obvious but equally true that *all* knowledge workers interpret and interact with the records of the past.

Along both vectors, the aim is to develop new insights and practices, to understand and advance what has been done, and to put energy into unduplicated effort. In short, to make work both smarter and more efficient. To do this, we need good access to both past and contemporary research.

If the Senate, its members and staff, and its records were scattered across the nation, the impediments to efficient work practice would be overwhelming. This is in some sense the state of the research community and its records, which are of necessity and for some very good reasons scattered across the nation. Furthermore, in distributed manufacturing organizations, the problem of coordinating the knowledge base and preventing the duplication of effort in large product teams is becoming increasingly difficult. Much time is wasted in rediscovering what someone else is doing or has done or in trying to find what related work is available. Practical, efficient, high-performance networks linking workers with each other could dramatically increase the productivity of all the different sites. The problems of separation might be overcome, while all the advantages of broad geographical distribution would remain.

2. Beyond search and retrieval --- interpretive synthesis

In finding precedents for their arguments, lawyers, judges, and scholars know they are not merely undertaking a task of search and retrieval. They are actively producing an interpretive synthesis out of their sources. This process is not reserved for lawyers, judges, and scholars alone. All knowledge workers, in manufacturing, in marketing, in sales, in customer service, now find themselves involved in complex processes, making sense of and making judgements about emerging and incomplete data. All face the demands of both producing and acting upon interpretive syntheses. As the amount of data they must work with continues to increase dramatically, they must be helped sieve through the haystacks of inconsequentiality to reach that important needle.

Donald Schön of the Massachusetts Institute of Technology has described this process as "seeing as" :

> When a practitioner makes sense of a situation he perceives to be unique, he sees it as something already present in his repertoire. To see *this* site [Schön's example here is drawn from architecture] as *that* one is not to subsume the first under a familiar category or rule. It is, rather, to see the unfamiliar situation as both similar to and different from the familiar one, without at first being able to say similar or different with respect to what. [*The Reflective Practitioner*, 1979, p. 67]

While data may sometimes be readily found, its interpretation and use are acts of "seeing as" in which the ability to separate what is relevant from what is not is crucial. Such decisions are reached through a process of judicious interconnecting, interaction, and interweaving of fragments of information.

The challenge of very large, linked databases is therefore not simply one of size alone. As information systems continue to expand almost exponentially, navigation through the data is of course a problem, but knowing what to look for and what to do with what is found are at heart of the challenge. Navigational aids will need to offer capacities beyond search and retrieval to be useful. They will have to enable people to get a sight of the whole forest as well as the individual trees --- to empower them to "see as," to support their interactive sense making, and to allow them to cope with, in order to capitalize upon, the data they are receiving.

Getting from raw data to a wise insight is a complex process that tempts us to add to T.S. Eliot's couplet

> Where is the Wisdom we have lost in the Knowledge?
> Where is the Knowledge we have lost in the Information?

the extra question

> Where is the Information we have lost in the Data?

With supercomputers capable of churning out masses of data, the prospect of getting lost is very real.

Information from data
Extracting information from data involves the ability to find or induce recurrent patterns. As we scan reams of data, implicitly what we are often looking for are patterns that match or resemble what we already know. In this process we are giving shape to or putting an interpretation on the raw data before us.

To match this sort of implicit process of inference, transformation, and selection, we can envisage knowledge robots as the next generation of smart navigational aids. "Knowbots" would first traverse the data, testing established patterns against the emerging stream until a set of adequate matches is produced in a process of searching "key patterns." Visualizing searches are particularly important because of the uncanny ability of the human eye to detect structural patterns in a visual field --- hence the power of "seeing as." With better visual selection and presentation of data, this ability can be more readily put to use. One of the original motives for building a high-performance network was, after all, to provide users of supercomputers in remote sites with opportunities to see structure in the flow of data from a visual display.

Knowledge from information

From the crude pattern matching of the first pass, the knowbot could reduce raw data to relevant information. In the next pass, it might be able to interpret explanatory hypotheses (in the spirit of analysis) that link information. The knowbot could search the information for instances that confirm or offer counter examples of the relevant hypothesis. Having first found the information, the next stage refines and enriches the thesis it supports.

In doing this, the knowbot, like a librarian developing a better understanding of what a reader is looking for, becomes a repository of accumulated and specialized knowledge. As a result, the interactions of the knowledge worker and the knowbot become involved in a bootstrapping process, each developing a continually more powerful ability to make sense of the data.

In this stage, the knowbot searches not so much for information as for the links that bind information into hypotheses. This also gives rise to the prospect of electronic serendipity, whereby the search tool actually produces connections that would not otherwise have been made. This is a process of automated theory formation that can produce new theories for verification.

Wisdom from knowledge

The final step involves understanding relevant theses suitable for application in a particular case. The knowbot now functions like an efficient, personalized reference assistant, who can find significant cases that permit arguments to be made from past events to future ones, enabling researchers to see what part of history might be applied to current conditions, to recognize how one instance provides adequate precedent for another. This movement from the development to the deployment of a thesis is a significant characteristic of expertise. It is not normally believed that technology can be directly effective in the development of wisdom; it is possible, however, to use technology to provide aids to help advance that development.

This sort of interpretation through a process of continual refinement, because it feeds off the assumptions underlying an argument is particularly effective at exposing those assumptions. This facility becomes critical in helping knowledge workers refine their productive interpretations and make insightful judgement calls.

CURRENT RESEARCH

Given the two central concerns outlined above --- supporting collaboration and aiding interpretation --- research at Xerox PARC has in part focused on exploiting and enlarging intelligent access to and collaboration with documents. This is evident both in the Corporation's *product* families and in the *processes* with which it designs, delivers, and maintains its products. Much of this work is proprietary and still under development, therefore here we shall be sketching out the areas of interest rather than the detailed results.

1. The documents of the future

Xerox is currently conducting research into the sorts of documents that will traverse information highways of the future and the sort of tools that will be needed to work productively with them. They are thus heavily dependent on the networks and infrastructure that will handle documents while they are simultaneously pushing at the frontier of advances in their design and use.

We begin with the premise that documents, broadly defined, are the substrate of collaborative work and innovation. Without documents to construct, construe, examine, pass around, alter, and amend, our culture could not have advanced to the stage it has reached and would not be able to continue to advance. But we are evidently moving into a new era of documents. And our aim is to move beyond reproduction of inert documents and beyond the handling of images to the creation, storage, and retrieval of "smart" documents, supported by an array of intelligent document services. Documents now, for instance, convey more and more images. They are also becoming increasingly interactive. The richer (and smarter) they get, the more room they need on any document highways.

The implications of such a trend are very serious for any national infrastructure. For people, a picture is worth a thousand words. But for a computer, in terms of storage, a simple image can occupy the space of a 50,000 words. Consider a standard page of 500 words. It requires about 2500 bytes or 20,000 bits. An image at fax quality, however, requires about 40,000 bits per square inch. At about 100 square inches to a page that's approximately 4 million bits per page. 16 levels of gray bring this to about 16 million bits, which when compressed by a factor of 10 still leaves us with 1.6 million bits compared to our 10,000 bits for a page of text. And this is without considering color or anything so ambitious as animation or video streams.

The conclusion of this abstruse calculation is that data highways that now just suffice for passing text around are going to collapse under the demand of the image rich, smart documents of the future. Copiers and sophisticated scanners process text at about 60 pages per minute. It is easy to see how documents only slightly more complex could quickly saturate existing national networks. Yet it is increasingly complex documents that offer the greatest capacity for effective communication and innovation.

2. Tools for the documents of the future

It is a mistake to believe that when people are joined by complex documents, work will continue just as before. The telephone, the photocopier, the electronic typewriter, and the desktop workstation helped to change office practices significantly and were themselves changed in the process. As complex document-processing capabilities are introduced, this reciprocal process will continue, throwing up new demands and new functionalities. At Xerox PARC, we have been exploring several instances and dimensions of this. Among others, we have focused on:

Media Space --- product and process

A powerful array of tools is necessary to facilitate the collaboration and distribution of work practices within the structures typical of decentralized, multinational corporations and between such organizations and their web of smaller suppliers.

But these tools may not just enable collaboration. They can, for instance, capture previously uncollected but valuable information. Some will, for instance, embody not just the product but the process of the evolution of an idea. And for many of us the process is essential to a full understanding of the end result. It is also vital for learning from errors and disambiguating results.

In one of the *Media Space* studies at Xerox PARC the intention was to explore the interactions and the needs of designers (three architects, in this instance) working on a complex design (an office space). Each was unknown to the other and all three were separated from each other. Their work was mediated by a computational/video medium. The system allowed them naturally to capture their developing thought patterns. Not only were the architects able to reach a rich collaborative design remarkably quickly, despite having never worked together or in this way before, but when it came to presenting their design to their client, they were able to use what the mediating devices had recorded to reveal the informative and complex evolution of the design, thereby rendering the truly innovative end product more understandable.

One essential element of this computer mediated, network intensive distributed work system was its ability to produce a sense of co-presence (much richer than teleconferencing) that simultaneously provided without extra overhead a means to capture experience for feeding and structuring organizational memory.

Information access --- needles and haystacks

In these documents, searching, as we have suggested, will be very different. The richer the documents and the larger the database, the more complex the search. It may be very much like the search for a needle in a haystack, except that we often begin the search not only faced with the haystack, but also not quite knowing that it is a needle we want out of it. Then the sort of symbiotic pattern matching we described above becomes increasingly important.

Early on in a complex search, in particular, the idea of what is wanted is often only partially coherent. But a request can be refined through a series of successive iterations. This calls for tools that can begin with indexical examples, model their properties, and through a series of iterations, retrievals, and refinements reach a more articulate and more accurate understanding of the query and ultimately, of course, an answer. It is thus important that national networks have the band-width to support these kinds of iterative searches.

Work practices --- the role of informal settings

Informal work practices are little understood and much under-appreciated. The casual chat around the coffee pot, for example, can not only be highly informative; it can also provide relaxed and therefore spontaneous and creative circumstances that are hard to duplicate in formal surroundings. Our study with the media-space voice and video link from Palo Alto, California to Portland, Oregon was an attempt to understand the informal practices of workplace communication and to discover how they are helped or inhibited with complex, high-bandwidth links. It is clearly wrong to believe that just because people can see and hear each other they will be able to interact spontaneously and creatively as they once did in informal settings. New theories of organizational behavior are just beginning to realize how important informal narratives and exchanges are for the production and dissemination of organizational knowledge. Systems that can support them thus offer a resource for organizational "bootstrapping," that can remain informal while making a vital contribution to the organization.

These are the sort of informative studies that help reveal what the properties and needs of a high-bandwidth network will be.

CONCLUSION

The aim of this testimony has been to throw enthusiastic support behind the concept and the potential of the National Research and Education Network. In conclusion, it is important to note that in order to fulfill that potential, we must not lose track of one central factor. We have argued that the abilities to meet present and future challenges of information technology are latent within this exciting prospect. The solution to these challenges is not, it must be emphasized however, wholly technological. In order to bring out the beneficial potential latent within a high-performance network, it is essential to understand and leverage the complexity and the richness of ordinary work practices. In our urge for technological advancement, these can too easily be overlooked.

Higher-quality tools are not necessarily the result of increasingly higher bandwidths. The "Post-It" has immense functionality, but very low technology. Moreover, increasingly higher-bandwidths often serves to make technology more complex and burdensome. As we build more powerful work tools, we must make the linking of people and machines symbiotic and synergistic rather than independent and potentially antagonistic. And to do this, it is necessary to understand the nature and the subtleties of successful work practices and to shape new work practices accordingly.

In particular, machines need to reflect their users closely. For instance, each community of experts traffics in distinct vocabulary and sets of mental models. Experts in a particular subject matter frame their hypotheses in their terms, which tend to be specialized and restricted but highly resourceful. Thus it may be less important to capture the breadth of "natural language" than to be able to reflect the language of the specific community of users in the system itself.

To make high-performance computers readily available to, for example, designers, engineers, and so forth, major advances in software will be required. For major advances, these cannot be merely libraries of canned procedures or subroutines. They must be rich, flexible, specialized programming languages and environments that directly reify the abstractions of their users or the subject matter. This is, of course, only a single example of the way in which machines should reflect the rich practices of the people for whom they are designed.

Finally, to add one small caveat, we should all remain aware of the national need to explore the vast spaces of novel architectures for special-purpose supercomputers. Large, general-purpose computers are unquestionably important. But a significant amount of capacity is devoted to maintaining their generality. It may often be more efficient to use special-purpose, dedicated processors whose architectures more naturally mirror the structure of the problem space to be explored for specific tasks rather than always to resort to the power of general-purpose ones. To this end, we must learn to think expansively about nonstandard architectures and approaches to high-performance computing.

Senator **Gore**. I really like that statement. Thank you.

We are going to go next to Dr. Irving Wladawsky-Berger, who is with IBM.

I appreciate the chance to work with you in the past in thinking through some of these things, and thanks for joining us here today. Please proceed.

STATEMENT OF DR. IRVING WLADAWSKY, DSD VICE PRESIDENT-SITE GENERAL MANAGER, INTERNATIONAL BUSINESS MACHINES CORP.

Dr. **Wladawsky**. Thank you, Senator Gore, Senator Robb. Let me just tell you briefly what I do. I am a Vice President in IBM's Data Systems Division, which is the division responsible for large systems development and manufacture.

I have direct responsibility for our high end operating system software as well as major portions of our super computing efforts. I am also a member, as is John Seely Brown, of the Computer Sciences Board of the National Research Council. So the subjects we are discussing here are really close to my heart, let alone my overall professional interest.

In the interest of time, let me go straight to try to answer some of the questions that you asked me to focus on in the testimony, starting with the question of what are some of the potential benefits of an information infrastructure.

As we all know, technology is advancing at an extremely fast rate. We have a lot of great technology out there in computing power, in storage technology power and in communications power. As we look into the future, it will only come at us faster and faster and faster.

I view the key benefit of the High Performance Computing Act as giving us the opportunity to take advantage of these technologies and translate them into leading edge applications that benefit our country.

In other words, by building a common infrastructure, a whole variety of people can build applications on top of that, and we enable those solutions to happen. People can concentrate on solving problems—namely, the applications—rather than having to concentrate on building the basic infrastructure.

Imagine where we would be in our transportation system if every time we left home we had to worry about whether the roads are paved, whether the gasoline stations are there, whether there is a place to have lunch. I do not think we would take as many trips as we do, and I do not think we would go very far.

A lot of building applications, especially advanced applications, is in that state. A lot of people have to do a lot more than they would like because there is a lot of infrastructure missing.

Now what are some of the examples of the wonderful applications that we can see the infrastructure making possible? We have discussed many of them throughout the morning. Let me just quickly go over some of them.

Some are applications designed to keep us at the very leading edge of science and engineering, and the recent OSTP reports, the latest one of which came out last week, cite them very well: mapping the human genome or advanced materials so we can build very light airplanes and so on. Those are wonderful applications. We should build them, and we fully support those efforts.

I think, however, that as a Nation we have to strive not just to be at the leading edge in developing new technologies but we should also try to transfer the benefits of high performance computing technology to as many people as possible, doing as many applications as possible. In fact, that is where I see the benefits of the High Performance Computing Act. How can we get applications to happen that affect our country's competitiveness and our quality of life?

We in IBM are doing a lot for our own interest to make sure that as many people as possible can build applications so that they buy our computers. Other vendors do the same thing.

The NSF centers are doing a remarkable job in expanding the use of high performance computing, and a number of state universities are doing similar jobs as well. More and more state universities have now set up super computer centers to help their research programs as well as state industries.

Let me just talk about one really interesting effort that addresses a question Senator Robb asked before. In South Carolina, Paul Huray, who used to be in Washington until recently, is trying to build a technology center linking the University of South Carolina, Clemson University, the technical college system of the state and private industry so that you can take the terrific skills in our universities, two-year colleges, four-year colleges, graduate schools and now make them available not just to large firms but to small and intermediate firms, tool and die manufacturers, people designing screws, whatever you have to do.

The states by following the leadership of the NSF centers, are now building their own networks in the states, and the tool and die manufacturer will know that by being hooked up in this state they are also hooked up into the whole Nation's infrastructure because in the center of the state their super computer will be hooked up to the follow-ons of the NSF net.

Scientific and engineering applications are very important. We also saw a wonderful example which John talked about in libraries and what it means to have online documents and what it means to have National Institute of Health medical databases and geophysical databases available to everybody.

Now with the massive computational power available and with the massive storage available and with the networks that we can put in place now, we are really in the position to give access to the best facilities in the country to just about everybody.

Any physician in whatever small, remote part of the country should be able to have access to the NIH databases so that they can understand what is going on, and if they can get their magnetic resonance images—we saw the power of that—and ship that around and get it to the right expert or the right super computing application, they can do superb diagnosis.

Senator Gore, you mentioned education before. I must tell you. I have small children, and I have watched them interact with Super Mario. It is clear that something fascinating is going on.

Senator GORE. Super Mario Brothers 2.

Dr. **Wladawsky**. They do both 1 and 2. They started with 1, and we had to go out and buy 2.

What is clear is that there is a marriage of the technologies of learning and the technologies of entertainment that will enable us to have children not just being educated but enjoying, being seduced by it instead of dropping out of school, which is what too many have been doing.

What is fascinating is that a lot of the same technologies for computer animation, which is what would enable a lot of the exciting learning tools, are the technologies that you would use in computer visualization, which is what you would use for computer fluid dynamic simulation.

So some of the things can come together, and it is a matter of getting the right applications.

Let me just finish by encouraging us not just to focus on the grand challenge applications. As a country, we always get excited by who hit the ball the farthest. We really want to swing out over the fences, but we really have to ask ourselves a few other extremely important questions like how do we design higher quality cars, higher quality screws from tool and die manufacturers, how can we get medicines to market faster, how can we put information technologies to use in educating our children.

I truly view the High Performance Computing Act as giving us a really effective way to now take on and answer those questions.

Thank you.

[The statement and questions and answers follow:]

STATEMENT OF IRVING WLADAWSKY-BERGER
BEFORE THE
SUBCOMMITTEE ON SCIENCE, TECHNOLOGY AND SPACE
SENATE OF THE UNITED STATES
SEPTEMBER 15, 1989

Thank you, Mr. Chairman. My name is Irving Wladawsky-Berger. I am a vice president of IBM's Data Systems Division, the part of our company responsible for large systems development and manufacturing. I have direct responsibility for IBM's operating systems software for large systems, as well as major portions of our supercomputing efforts. I am a member of the Computer Sciences and Technology Board of the National Research Council. So, the areas covered in the High Performance-Computing Act are very close to me.

As we all know, information systems technologies have been advancing at a very rapid pace, and will continue to do so well into the future. This is true in the actual computing capabilities of the largest systems as well as in small systems which bring considerable power to the desktop. It is true in technologies that store ever increasing amounts of information, permitting people to access large data bases and libraries that increasingly contain not only traditional data -- or words and numbers -- but also images, or pictures, including full motion video images. It is also true in communication technology, where information is transmitted at much higher speeds than previously possible.

The High Performance Computing Act of 1989 offers the opportunity to take advantage of these advances in technology and translate them into leading-edge applications that will benefit our country in a variety of ways. By concentrating our efforts on applications, we will provide solutions to some of our country's most critical problems. And in so doing, we will advance the state of the art of high performance computing by stretching the technology to meet the application challenges.

To date, much of the work in high performance computing has focused on advanced scientific and engineering applications. The number of users of supercomputers has significantly increased in the last few years, primarily in our larger universities, industries and government laboratories. Exciting "grand challenge" applications have been identified in the recent reports of the Office of Science and Technology Policy (OSTP). These are very important efforts designed to keep our country at the leading edge of supercomputing technology and applications. We support these efforts, and the objectives of the OSTP Federal High Performance Computing Program.

But there is an additional, very important direction that we must pursue, one in which this Act can take a lead role. As a nation, we must strive not only to keep high performance computing technology at the leading edge, but we must transfer the benefits of high performance computing technology to a wider set of people performing a variety of more fundamental applications directly affecting our country's competitiveness. Let me talk about some examples.

Hoping to encourage the growth of applications, IBM has established two Numerically Intensive Computing Centers in the United States that serve to assist customers in this area and help IBM research and development learn from these application requirements. The Centers are located in Kingston, New York and Palo Alto, California. Application areas supported include structural analysis, fluid dynamics, computational chemistry and physics, reservoir simulation, seismic data processing, circuit design, nuclear engineering, atmospheric sciences, and financial analysis. Each of these Centers establishes joint activities with engineering and scientific organizations in both the industrial and academic communities in order to expand the number of existing applications for large IBM systems and to advance high performance computing generally.

IBM is certainly not alone. Similar programs to expand the use of supercomputing are being undertaken by other vendors, the National Science Foundation Supercomputing Centers, and a growing number of universities. In fact, a number of states are establishing supercomputer centers to assist in economic development and strengthen scientific research. One noteworthy example in South Carolina will link the University of South Carolina, Clemson University, the Technical College System of the State, and private industry to bring world-class high performance computing applications to small and medium sized companies throughout the region.

Because the skills needed to develop such a system are significant, they will be contained in the university. These high performance computing experts will develop applications for use by industry employees with a minimum of computer knowledge and ability. The State hopes to improve the productivity and competitiveness of its existing companies and, it is anticipated, this service will offer significant advantages in drawing new companies into the state. This is a model that needs to be encouraged and replicated.

Replication and evolution of efforts such as this, however, will require focus not only on increasing the speed of computation but on advancing technologies for storage and retrieval of information, including storage of images, sounds and pictures. Such advances are critical both to scientific and engineering applications, as well as to many other advanced applications that manipulate vast quantities of data. Two examples include computer access to the contents of the Library of Congress or the medical data bases of the National Institutes of Health. Additional applications, however, are limitless.

With such massive computational power and information available, the challenge is to establish an infrastructure to give access to as many people as possible. Senator Gore, this is the idea behind your "National Data Highway", an interstate system for the transport of one of the most important products of the 21st century and beyond -- information. Technology is aiding the creation of such a network. In fact, only two weeks ago IBM demonstrated a new semiconductor chip able to transmit and receive data over fiber optic lines at speeds of a billion bits a second. These new optoelectronic chips are 50 times denser than earlier components, and contain 8,000 specialized transistors on a device the size of a human fingernail.

Imagine the capabilities of such an integrated system combining powerful computing, limitless data bases, and a comprehensive network. It could help scientists in their struggle to better understand the structure of human cells in an effort to win the battle against AIDS or Alzheimer's disease or multiple sclerosis or cancer. Access to such capabilities would help them model cells and simulate disease processes in a computer where their evolution could take place in minutes rather than months or years in the laboratory. Further, consider the benefit to physicians throughout the country who would have the capability to transmit medical images and data in real time to specialists -- either a human expert, or an advanced medical computer application -- for rapid, sophisticated analysis irrespective of the patient's geographic location.

Consider the impact such a system could have in helping solve the education crisis facing our country. Such facilities, bringing new creativity to education with techniques such as visualization and interactive computer animation, could be a significant aid to teachers in their challenging work.

Imagine young children carrying on a computer generated and controlled dialogue with animated characters who would not only entertain them but teach them as well. As the children get older, imagine incredible, visual simulations of world events, chemical processes, biological problems, or literature which the student would not only see, but interact with. As they move onto higher education, think about the benefit of students and researchers having access to the Library of Congress through a high performance computer network. The potential is enormous.

All of this and more would be possible with high performance computers and sophisticated networks. But if we are to realize such useful applications for high performance computing, we must move beyond our traditional approaches which have tended to focus on developing ever more advanced technology rather than focusing on its application.

Sometimes we are obsessed with hitting home runs rather than concentrating on the fundamentals that score runs and win games. Our focus is on whose machine is the fastest, whose "grand challenges" are grandest, who gets the most prestigious award. While all of this is important, we must not lose sight of the other -- perhaps less glamorous, but possibly more important --

questions: How do we design higher quality cars; How do we bring medicines to market more quickly; How can we apply information technologies to assist in the education of our children? Answers to these questions will come by focusing on designing high performance computers that exploit the advances in technology AND by focusing on applications -- that is, the software that instructs the computer to perform specific tasks and solve specific problems.

To achieve these objectives, significant work will be required in the hardware, software and networking areas. Work the private sector does best should be left to the private sector, including hardware and software design and manufacture, and the operation of the highly complex network required to communicate between systems and to exchange information. Our government, however, can be of tremendous assistance by focusing attention on the use of high performance computing to strengthen both American competitiveness and scientific advance; by serving as a catalyst, using high performance computing to advance national interests; and by participating with the private sector in funding programs and transferring critical skills.

The National Science Foundation Supercomputer Center approach is a very good start, one that has served as a catalyst to efforts by state governments and universities. We must make sure that these Centers remain strong. Programs to improve access to these facilities by small and medium sized companies -- such as the effort in South Carolina -- would provide a tool that could strengthen these firms' competitive posture at a price they could afford. Finally, government assistance in the construction of a nationwide network, built with expansion potential in mind, could make a profound contribution to a truly national high performance computing capability that would benefit us all.

In summary, we support the objectives of the High Performance Computing Act of 1989, though we do have some concerns involving standards and intellectual property aspects of the bill. We encourage a national plan with an application focus supported by a comprehensive network. We believe the time for action is now, and we are persuaded that work resulting from implementation of the High Performance Computing Act will encourage governments and the private sector to build advanced applications faster than they otherwise would. We are pleased to work with Congress toward implementation of such a plan.

QUESTIONS OF SENATOR GORE AND THE ANSWERS THERETO

STANDARDS

Q1: What are the principal issues in ensuring that users can effectively and conveniently access and integrate information from a variety of databases and a variety of networks?

A1: The principal issue in ensuring that users can access and integrate information from multiple sources is to ensure that open interface or protocol standards are available as necessary to provide a common basis for the meaningful exchange of information. This does not preclude the use of proprietary approaches in either databases or networks since the open interface or protocol standard becomes the access and interchange mechanism. Fundamentally this has been the purpose of the development of the Open System Interconnection (OSI) and office system standards which have been developed in recent years as an aid to facilitating the exchange of information in the Information Technology industry.

Q2: Will standardization create undesirable side effects? What can be done to reduce them?

A2: The development of standards often involves the selection of one technical approach or specification over another. The proponents of approaches not selected as the standard may feel disadvantaged since their choice was not accepted. However, experience has shown that when standards are developed in an open voluntary consensus process in which all interested parties have an opportunity to participate, the resulting standards that are developed provide significant benefits to users, manufacturers and others.

Q3: Who should be in charge of developing the necessary standards?

A3: There already exists within the information technology (IT) industry a standards development mechanism which has for many years proven its ability to develop standards that are both technically and economically useful to manufacturers and users. This work takes place in the ISO/IEC Joint Technical Committee 1 (JTC/1). The JTC/1 committee has the primary IT industry responsibility for the development of international standards across a wide variety of IT subjects including data formats, database standards, telecommunications, programming languages, etc. The ISO/IEC standardization work is conducted on an international basis to further ensure worldwide agreement. It is a voluntary consensus based process with strong and active participation by users, manufacturers, governments and other interested parties. The ISO/IEC JTC/1 committee is and should remain the primary organization for the development of IT standards.

NETWORK SECURITY

Q1: How serious do you think computer viruses are? Do you think that technological fixes exist?

A1: Computer viruses are a relatively new problem to the computer industry. No system is immune to harmful code. The number of known incidents, although relatively small, appears to be increasing, and is thus becoming more serious. Today's defenses against viruses rely on effective security practices by users. These practices include making frequent back-up copies of important data, controlling the source of software, and using virus detection programs that are available from IBM and other vendors. In net, we rely on diligence and technology for protection.

Technology does exist that is beginning to be used to enhance several aspects of computer security, including virus prevention. Cryptography has long been used to prevent disclosure of sensitive data. More recently, led by the banking community, cryptography has been applied to prevent data or communications from unauthorized modification.

IBM offers an array of products to protect systems from unauthorized access, and new technological measures are being sought to further address this new threat of viruses. In addition, IBM continues to emphasize the importance of good general security practices, management controls and user education for safeguarding data and preventing the introduction of viruses.

Q2: Could concerns about computer security prevent the widespread adoption of computer networks?

A2: Computer security concerns need to be evaluated balancing the risk of viral infection or security breaches against the benefits of an open network enabling the rapid flow of time-sensitive information among users. Good general security practices will reduce the risk of accidental viral infection, and anti-virus practices will aid in the prevention, detection and spread of a virus. Effective security software and consistent applications and procedures will help to minimize some of the risks. Thus, while these concerns must not be ignored, IBM believes the far reaching benefits society derives from the use of computer networks will result in their continued and increased use. Security, however, is an important requirement for widespread adoption of computer networks. Two key requirements are:

1) positive identification of users accessing the network and;
2) controls over what those users can do on the network.

These requirements can be met with today's technologies, but not all systems and manufacturers' implementations employ them, nor are system procedures across current research and education networks structured to require them. User identification can be validated via passwords or more sophisticated mechanisms employing "smart cards" containing electronic circuitry embedded on a credit card sized token or various biometric mechanisms, such as electronic signature verification. Today's network protocols allow for security functions to authenticate the identity and source of requests, validating the successful transmission of the communication, and authenticating the contents of the delivered message. As noted in response to question 1, such mechanisms are most often implemented with cryptographic algorithms.

Q3: A particularly frightening thought is that the databases we are talking about will be penetrated and altered. A database is not very useful if users cannot be certain that the data in it is correct. Are there ways to ensure that databases have not been altered?

A3: Yes, many database products contain security control mechanisms based on authorizing particular users to specific portions of the database. The security process, therefore, is to ensure positive identification of the requesting user, and to conduct the administrative task of specifying the rules and rights of access for users. The size of this task depends on how complex an authorization structure is desired by the data owner. It is trivial to set up rules such as: anyone can read the data, but only a specified list of individuals are allowed to modify the data. More complex rules would be required to let various individuals have update rights to different portions of data. But this is an administrative task and does not require new database technology.

Q4: I would think that there would be a lot of money to be made in computer security. Dr. Waladawsky-Berger, what is IBM doing in this area?

A4: IBM's commitment to data security is a long standing tradition. We have had a corporate data security policy since 1973, under which we ensure data security is a basic design criterion for future systems and products.

IBM continues to work extensively in many areas of computer security by marketing security products, but more importantly by upgrading the system security functions in the processors, operating systems, database managers, and communications products that make up a total system. We recently (October 24) announced our directions in computer security for a broad range of computer users. Key items announce include new more secure operating systems, meeting the Department of Defense Trusted System Criteria, a new family of cryptographic products for enciphering data transmissions, message authentication, and user identification; security consulting services to assist customers in assessing risks and formulating control policies; and guidance to our customers regarding anti-virus measures and detection programs. We believe proper security controls are a requirement for continued growth of the computer industry and we are working to provide those solutions across our products.

DATA TOOLS

Q1: Do researchers have the tools necessary to cope with all that information?

Q2: What kind of software and hardware are your companies developing to deal with all this data?

A1&2:

Data base technology has made large strides in the past 10 years both in the software used to create and retrieve the information and the cost per bit of storing the information.

However, researchers today do not have many of the tools necessary at their disposal to easily create terrabyte databases for specific disciplines and easily search the information in the database in the languages of the researchers. This is a complex problem that grows in complexity when data such as Earth Resources Data is stored in differently structured databases in different physical centers with different access policies, etc.

The Federal Government can help here by establishing centers that will organize important research information, and by working with vendors of hardware and software, thus ensuring that the data structures and query systems lend themselves to efficient retrieval in the language of the researchers.

IBM will continue to improve its cost per byte of storing information and the retrieval systems needed for both hierarchical and relational data. However, more work is needed and can be accomplished via government sponsored centers focused on discipline specific data such as DNA or molecular structures. Here, knowledge based systems, which we are also working on, can be valuable aides in the creation and retrieval process.

In addition to the on-line or immediate access part of the databases, there is an important area that IBM is working on relating to backup and archival systems. The key idea behind these systems is that infrequently used data is moved to a more cost effective media along with an increase in time to retrieve it. This provides a good balance between cost and performance and is critical to very large databases.

Q3: In five years, what will I be able to do with my personal computer that I can't do now? What kind of data will be at my fingertips and what will I be able to do with it?

Q4: What about in 10 years?

Q5: What can the federal government do to speed along these developments?

A3-5:

In general, one has only to look at the types of power and data that are available on today's mainframes to see the kinds of power that will be available in a personal computer in the next 5 years. This power will be used to allow researchers to communicate with the personal computer using a more direct form of interaction. An experimental workstation interface, jointly developed with the laboratory for Solid State Physics at Cornell, is one of the steps being taken in this direction. This interface will allow physics researchers to point and click on a periodic table to bring up models of the individual elements that can be combined to form new solid state compounds.

The Federal Government can help by supporting the researchers in universities who work on projects with industry and government to make the strides that help make researchers more productive and private industry more competitive.

NETWORKING AND DATABASES IN OTHER COUNTRIES

Q1: Have its potential strengths or weaknesses been evaluated? What do you think of Minitel?

Q2: What can the U.S. learn from the French that would help us to tackle our evolving systems?

A1&2:

Minitel has had mixed success. Although the growth figures for Minitel terminals have continued to increase, so have the give-aways of these terminals to households and businesses by the French Telecom. In a recent study authored by Gilles Ghesquiere and co-published by Samara Associates, Cambridge Massachusetts and Communications Trends Inc., Larchmont, N.Y., over 50% of the Minitel households said they would return their terminals if the French Telecom charged approximately $2.00/month. In addition, the growth rate for average usage/terminal has been declining and was lower in 1988 than 1987. Moreover, heavy users declined 7% in the three years ending in 1988. One of the reasons for this decline in usage of Minitel is the provision of free terminals. Minitel provides free terminals placing them in the hands of users that have little interest, ability or resources to use the service.

Another point that is critical in viewing future services such as this is the overall system design. The computing paradigm has changed since the French did their initial study in the 1970's. At that time, there were no PCs in the home. Today in the US over 20% of the households have PC's. The resulting overall system design using dumb terminals for the end user places the computing load at a central site and requires a more significant and costly transport service.

The paradigm shift of PCs and intelligent intermediate nodes of today drives a different overall systems design. Pivotal here is the trade-off between computing at different stages in the network and overall telecommunications costs. The power of end user terminals or workstations will continue to increase well beyond the capability of the centralized host of the 1970's. The technology and cost trade-offs are very different with this distributed computing paradigm. This places computing power in the user's home and intermediate nodes in the network to help with the user interface, information requests and presentation of results. The hosts are still used to manage the large databases, but their role in the network is quite different.

The lesson the U.S. can learn from the French Minitel initiative is that whatever we do, our efforts should be structured to allow for future technological advances, for this new technology will yield the greatest overall cost benefits. Other important issues are fixed pricing for the services vs. metered pricing currently in Minitel, and issues of competition. Competition is key in the deployment of the service to permit the best market driven approach for the end users and data providers.

THE SOFTWARE GAP

Q1: How serious is this problem in the area of data management? Do we have the programs needed to handle the data that supercomputers can process?

A1: Many supercomputer applications are characterized by the need to process very large amounts of data such as for seismic exploration. These data are typically arranged in simple structures such as large arrays. The problem in supercomputing is usually either the complexity of the algorithm which is applied against these data and/or in the shear volume of data which have to be processed and stored. The challenge of data management in supercomputing is to handle input and output in such a way that the slow access to disks and tapes does not become the bottleneck for the overall application. Hardware improvements such as increased channel and access speed, and splitting the data across multiple channels and/or disk heads, will help to ease this particular problem. IBM is working in all of these areas. Of equal importance are the database systems and algorithms that take advantage of specific disciplinary databases and help the supercomputer users to create and easily retrieve the information in the language of the supercomputer user.

Q2: Is the gap between computer hardware and the computer software growing?

A2: Generally the progress in hardware technology has been much more rapid than in software engineering, which has had few significant breakthroughs since the invention of high level languages such as Fortran 30 years ago. There have been some quantitative improvements in the modularity of code functions, improved data structures and more disciplined flow control, but no fundamental breakthroughs for improving programmer productivity. Completely new approaches such as functional languages are being developed, but have not yet made significant inroads in the daily practice of data processing. As a result of this, the software gap is growing.

Q3: What can be done to reduce this gap and what role should the federal government play in this effort?

A3: Standards and open systems such as UNIX will ease re-useability and transportability of code and increase productivity. The federal government can help in encouraging the setting of standards through participation in open standard-setting committees.

Government support for research into highly parallel system software and algorithms would be helpful. Many of the applications currently run on today's supercomputers could benefit from these emerging hardware architectures. Another area where government support is essential involves research into new computing paradigms, such as the recent DARPA project for supporting neural network research. This research could also help in finding the really fresh new approaches needed to overcome the software gap.

Q4: I understand that many of the new massively-parallel supercomputers are particularly good at handling huge data sets. Why is that?

A4: Massively-parallel supercomputers often have extremely large real memories (up to several thousand megabytes) which is key to effective handling of very large data-sets. Combined with this is a very high peak performance (speed), which in many cases can actually be exploited for real application programs.

Q5: Does the software exist to properly exploit these new massively-parallel machines?

A5: The software for exploiting massively-parallel machines is in its early stages of development. On an absolute scale these computers are still harder to use than conventional machines, but the promise of nearly unlimited performance and very good price/performance are such powerful incentives that these machines will enter the mainstream of large-scale computing in the next few years. Some of the critical software components that need to be addressed are; compiling and debugging, as well as algorithms for many of the key application areas such as fluids, molecular modeling, structures, weather, etc.

NEW TECHNOLOGY

Q1: Do you expect the steady, rapid improvement of data storage devices to continue? Will new approaches be required, or can we just keep improving the present kinds of storage media?

A1: In the field of data storage technologies and products, I continue to see rapid advancement of technologies critical to the evolution of advanced storage devices. Conventional inductive magnetic recording technologies continue to lead the parade as higher and higher areal densities are achieved. The evolution to smaller hard disk diameters resulting in higher performance and higher reliability products has been a boom to all data processing systems designers and users. The hard disk business is conservatively estimated at $20B today worldwide and can easily be projected to double in the next 3-5 years.

Flexible media, such as magnetic coating on a flexible substrate like mylar continues to employ the latest technology improvements such as rotating heads from video recording type products. Both the hard disk and the magnetic tape have benefitted from advanced technologies such as improved magnetic particles, thin film coatings on hard disks, thin film magnetic transducers and the very critical low cost LSI circuitry required to perform the signal processing to enable high density magnetic recording embodiments. Continued government focus on magnetics research would be helpful.

There is an exciting new technology on the horizon being driven currently by the consumer market, the optical storage area. The CD-disk is currently in most households where hi-fidelity sound reproduction is desired. The current trends to utilize solid state laser read/write optical heads in combination with either a phase-change or a thin film magneto-optic material can produce areal bit densities approximately 5 times higher than convention inductive magnetic recording technologies. The introduction of these magneto-optic or phase change optical storage products has been slow to start due to their current low performance characteristics compared to that of the hard disk or the flexible tape products. Currently, their cost is high due to low volume demands. The capability to employ the removability and interchangeability of the optical disk is a significant advantage over the rigid disk storage product and this interchangeability, coupled with future technological advancement, will result in future optical storage products that could one day replace many rigid disk products currently used.

During the next ten years, it is clear that inductive magnetic recording (including magneto-resistive reading) and optical storage will be the major product technologies in contemporary dataprocessing installations. There remains considerable growth for these technological capabilities with current evolutionary progress to be the mainstay of data storage in the near term. It is quite possible that the technologies already in hand in this area could produce products to satisfy dataprocessing demand for the next ten years.

Q2: If data densities improved a thousand-fold, what kinds of applications would be possible?

A2: Assuming that the technological progress projected follows historical lines; i.e., performance parametrics also improve consistently with recording densities, but not necessarily at the same rate, then many of the current applications associated with imaging will be expanded and become common place. Since both major storage technologies are considered to be semi or permanent storage, then the projected future high densities are also considered to be semi or permanent storage. Such high efficiency storage capabilities would permit a major penetration of conventional paper and microfilm storage.

Automated digital storage techniques permit on-line accessibility as well as computer manipulation of large databases. It would entirely modernize existing business and governmental operations. Printed documents, although still in use, would become largely obsolete being replaced by work stations directly accessing large interactive databases. Consistent with high density storage technologies, data compression technology will be used extensively to enhance storage capabilities even further, cut down on the bulkiness of documents, drawings and picture storage requirements. The marriage of digitized data intermixed with digitized pictures, graphs and drawings will become common place and one could predict that interactive, high definition video could be stored in the same data structures.

I would not be stretching my imagination too far in predicting great changes in the work place if data processing capabilities improved by a factor of 100 - 1000X at roughly the same cost, i.e., a price/performance improvement of 100X. This would enable individuals, as well as business and the government, to have broad-scale access to on line databases with flexible data communication capabilities such that distributed databases, as well as large centralized databases, could flourish interchangeably. I could see interactive video being used extensively in the work place.

Q3: Can we predict what the limit is? Or will we one day be able to have the entire Library of Congress stored on a cartridge that would fit in a shirt pocket?

A3: From my experience, to predict a limit is nearly impossible since our vision is clouded by the technological progress that we can foresee. For the present let it suffice to say that I see no practical limit such as the magic 1000x over the next 10 years. Today it is not a hazardous prediction to envision terrabyte storage systems that could fit into a package the size of an average refrigerator within the next 2-5 years, at a price every medium to large size dataprocessing installation can afford and would install. By that same process your every day personal computer would have a gigabyte of rapidly accessible data via an optical disk drive.

As for one day being able to have the entire Library of Congress stored on a cartridge that would fit in a shirt pocket, that is possible, although highly unlikely. One of the major problems of the technology advances that I have targeted is that they require extreme precision to maintain reliability. Therefore cleanliness, and more importantly, environmental control are required. The major impediment to technological progress is removability which requires mechanical components and tolerances not amenable to high precision electro-mechanical storage elements. Although the servo-controlled access mechanisms for optical storage systems of today have demonstrated significant progress over those used in hard disk files, the fact still remains that the removability of the optical disk is a limit to the storage efficiency of that technology.

Just as in the case of the hard disk, totally enclosed (non-removable) optical storage elements will be a way of the future to achieve the efficient storage required to fulfill the very large, on-line storage systems of the future. So in net, while it is possible to predict the capability to store a database the approximate size of the Library of Congress on a cartridge (about the size of a 3.5 inch optical disk of today) that could fit into one's shirt pocket, I doubt that one could be made in the next 10 -15 years. Rather, one might predict that the actual data could be stored in a volume that is much less than a cubic foot of space, but the total electronics, environmental controls and power system could enlarge the package to the size of a small refrigerator.

I would not hazard a guess about what might happen in the period much beyond the year 2000. Some new technology or significant breakthrough in one of the existing technologies, could enable advancements yielding capabilities beyond one's perspective. Although there are proposals for three-dimensional storage arrays today, most suffer from a depreciation of another dimension; i.e., to gain volumetric efficiency one must give up reliability or performance. To be practical, a storage system of the future must yield enhancements in areal densities as well as improved reliability, improved performance and no increase in cost!!! Now that sounds like a big job, but history is an excellent teacher, and electromagnetic storage systems have delivered on these factors over the past 40 years. There is little evidence this progress will not continue well into the next century.

Senator GORE. So this can help us not only to sequence the human genome but also to make a better mousetrap or a better computer chip?

Dr. WLADAWSKY. Absolutely.

Senator GORE. Thank you.

Our third witness will be Mr. Richard T. Wood, Senior Vice President for Business Development at University Microfilms, which distributes journals on optical disks for use by libraries and businesses.

Please proceed.

STATEMENT OF RICHARD T. WOOD, SENIOR VICE PRESIDENT FOR BUSINESS DEVELOPMENT, UNIVERSITY MICROFILMS, INC.

Mr. WOOD. Thank you very much, Senator. We welcome the opportunity to comment.

As you mentioned, we launched several products this year which, for us, were innovative and for libraries as well. And the proof of our success was the fact that I went down to the business school library at the University of Michigan at 7:00 a.m. on Sunday and found students queued up to use the system. So, it was nice to see.

Several thoughts. One is the fact that I think the costs that we are contemplated to develop this system is probably much too low. I do not think that is a negative. I just think it is more realistic to think that the cost structure is going to be extremely high and that the ongoing costs are going to be extremely high as well.

I also would like to see as part of the development of the infrastructure that we consider building up the resources of the universities at the local level.

One of the things that we have found in working with universities is the fact that many of them have not adequately invested in the correct amount of computerization at the library level, and that is partially due to the way they fund the libraries and the fact that, for example, those funds come out of a capital budget rather than a materials budget.

I think the concept that you are contemplating requires additional focus. I think your choices are going to be so wide as to what you want to put on the network system that we are going to have to make priorities in deciding which goes first, and I would like to see the focus be on that.

If I can bet anything, I think the bet is going to be that you are going to have much more demand almost exponentially than you forecast today.

Senator GORE. May I interject there?

Mr. WOOD. Surely.

Senator GORE. I agree with you. The model I have in mind requires that we install some high-capacity network links and then, as a result of that exponentially increasing demand, one will be able to charge user fees and private incentive-driven companies, state and even local governments will build the equivalent of feeder lines, and then the expansion will come very rapidly because of that increasing demand.

I think the cost of the initial backbone network and the initial digital data library is, in fact, quite low. But the full use of it and the expansion of it, of course, will involve much more money. But, by that time, we will have a fuller private sector participation in this and much more available sources of money to finance it. I am sorry to interject that.

Mr. Wood. That is precisely the point I wanted to make.

I think in addition to that I want to emphasize that the private sector very much wants to participate in the development and in the use of the proposed systems, and I think speaking as a publisher and speaking as an information provider, we very much want to participate in a very positive way in the use of the system.

We hope that as the bill develops you will encourage the private sector involvement.

[The statement and questions and answers follow:]

STATEMENT OF RICHARD T. WOOD, SENIOR VICE PRESIDENT, BUSINESS DEVELOPMENT, UNIVERSITY MICROFILMS, INC., BELL & HOWELL INFORMATION COMPANY

University Microfilms, Inc. (UMI), a Bell & Howell information company, is pleased to have the opportunity to comment on S. 1067 and the need to provide Americans with greater access to information. In particular, I would like to provide insights from our own market experience, as a leading information provider.

Electronic technology is altering the way information is gathered, produced, published, distributed, accessed and retrieved. It is a time of change and challenge for authors, publishers, information centers and users alike. While technology, of itself, will not increase directly the sum of knowledge, it does make it more malleable and more accessible. It has been our experience that it expands the total universe of users and improves the quality of research.

The concept of a national information network is seductive. Much of the technology is either available or within reach. The question of how to fund, however, is a serious issue. More importantly, what information should be carried on these proposed electronic highways and is it worth the high level of investments? Also, what are the roles of the private and government sectors? I would like to reflect on these issues in the process of describing several innovative electronic information programs recently undertaken by UMI.

UMI APPLICATIONS

After a year long market test and many years of development, UMI released in January 1989, Business Periodicals Ondisc (BPO), a database of 125 CD-ROMs. BPO integrates fully searchable ASCII coded abstract and index data with the fulltext facsimile images of the pages of 325 business periodicals. Search and retrieval is done at an advanced, but standard personal computer. Search results and facsimile article pages are laser printed. Each transaction is recorded and is the basis for tracking what is printed by: publisher, journal, issue, article, and page. Royalties are paid to publishers accordingly. A publisher and a library for the first time, now can track intelligently what has been reproduced from a publication. The commercial success of BPO is not yet assured, although the response to date is most encouraging.

Also in 1989, UMI, The Institute of Electrical and Electronics Engineers, Inc. (IEEE), and The Institution of Electrical Engineers (IEE) jointly initiated a market test of a CD-ROM database of IEEE/IEE journals, conference proceedings and standards, linked to equivalent INSPEC index and abstract records. As in the BPO product, pages are laser printed and transactions recorded for tracking purposes. Considerable experience has been gained already from twelve diverse test sites within the U.S. and Britain. Product/market decisions will be made jointly later this Fall.

The CD-ROM format was chosen because of its relatively low edition costs and the availability of inexpensive disc drives. The favorable economics were achieved thanks to the success and scale of the compact disc music industry. However, UMI remains media independent and will continue to use the best and most economical distribution medium available that meets market needs. In any event, the amount of data being handled is not trivial. For example, the 600M pages (63 Gbytes of data) currently in the BPO product are on 125 CD-ROMs. The UMI/IEEE/IEE product contains an additional 20 discs or 10 Gbytes of data. When we consider the fact that there are approximately 200M periodicals published worldwide, one realizes that the universe is astonishingly large. Clearly hierarchical priorities need to be applied in selecting what deserves to be stored electronically.

We foresee the ability to receive ASCII fulltext tapes directly from publishers, using Standard Generalized Markup Language (SGML). The difficulties associated with integrating graphics in the page make-up are considerable and we expect that it will be three to five years hence, before direct publisher input is practical. UMI assisted in underwriting the SGML effort and believes that both image and ASCII fulltext data must be considered in any information system design.

After nearly two years of market experience, it is clear that users relish accessing information electronically. Two market imperatives have emerged: the need for networking and the need for large scale storage devices. UMI is working now to develop local network solutions, in order to bring workstation costs below $5M each. We are testing also a 270-disc jukebox that was made in Britain by Next Technology. Several large computer technology firms are also working with us to find generic solutions to these issues.

IMPACT ON PUBLISHERS

UMI has developed the first generation single-user workstation of large-scale local CD-ROM storage of electronically published information. There is a demonstrated demand for this type of primary information in the marketplace. The immediate access to source material, the ability to retrieve high-quality graphics, and the simple, straightforward interface for the use of the system have proven very interesting and useful. Acceptance of the system has been outstanding, with the main user disappointment being that not enough material is yet available in the database for complete searching of a topic.

The next major development step for UMI is to begin large-scale production of electronic versions of the technical information needed by researchers, in cooperation with the major publishers worldwide. As indicated, the very large volume of data this represents necessitates expanding the UMI system from single-user access to multi-user access.

Some assert that advances in technology will require a rewrite of the copyright law. To the contrary, UMI feels that the copyright law works well. UMI has been successful in securing the cooperation of publishers in developing the electronic products and is licensed by the individual publisher by contract. Since being founded fifty years ago, UMI has observed the copyright law consistently and has served as a useful bridge between publishers, libraries, and users alike.

Publishers are concerned rightly about the impact of technology. Photocomposition is already extensively used to both lower costs and create derivative products. Nevertheless, the economics of publishing are usually precariously balanced. A loss of income from one format can endanger the whole. Consequently, technology is approached cautiously. The printed book or journal remains a very low cost distribution medium. It will continue to be valued as the primary source, because it delivers information exceedingly well: graphics, photographs, and color included.

Our experience to date indicates that a plurality of distributed formats will coexist, for the foreseeable future. Format cannibalization is a major market issue for publishers and information providers, however.

RECOMMENDATIONS

In order to reach the researchers around the country, there needs to exist a high speed, high volume wide area network such as the proposed NREN. Additionally, a general upgrading of local computer equipment to handle high quality, high speed, and high volume transmissions of data locally and from a centralized source will be needed. Costs of creating national and local networks for centralized distribution of information will be large -- possibly too great for any one organization to effect alone. To achieve the network objectives cost effectively requires close cooperation with the private sector. Standardization will be necessary for linking researchers to information providers, such as UMI. Public policy must foster standardization and encourage private sector initiative and involvement.

The electronic networking of researchers within the United States will be an ultimate necessity for scientific and commercial competitive progress. In some instances, researchers have already built their own professional networks, albeit on a reduced scale. It is our recommendation that a national effort be undertaken to design and implement a high speed, high capacity network, but that commercial investments of telecommunication companies, computer technology firms, and information companies be elicited first. We also urge focusing on ways to encourage the development and retooling of academic and research institutions so that they can build local area networks and make better use of information already distributed electronically.

If the object of the proposed national network is to simply distribute already published materials, than the high investment is unjustified. If this information is worth distributing electronically, the private sector will do it.

The development of a national information infrastructure will permit access to more source information than through traditional library means. This will reduce the amount of time needed to access and transmit data necessary for governmental and business decision research and analysis. Publishers may be more willing to allow electronic versions of their source publications with the wider distribution of the electronic network. The pinpointing of information needs would potentially reduce the waste of collecting too much general material. Ultimately, it could reduce the overhead of archiving multiple sources of decaying published materials.

The cost of such a national system will be enormously high. Usage cannot be free and ongoing maintenance costs will be substantial. Because no economical model exists, we are not yet convinced that centralized distribution is a viable cost alternative to distributed information databases. Government should not become, in our opinion, the chief information producer or provider.

For our part, UMI will continue its development of a flexible interface that would permit many types of software users to access the UMI resources.

QUESTIONS OF SENATOR GORE AND THE ANSWERS THERETO

Q-1. STANDARDS

We have already heard quite a bit about the need for standards. Progress is being made to create computer networking standards, although much more will need to be done before the NREN is built. The situation is worse when it comes to data formats and database software. There are a bewildering array of different ways to store and access data. If electronic documents are going to replace paper, we need to have some agreement on how they will be represented in a computer. Work needs to be done to ensure that users can easily access different systems and transfer data and software between different systems.

- What are the principal issues in ensuring that users can effectively and conveniently access and integrate information from a variety of databases and a variety of networks?
- Will standardization create undesirable side effects? What can be done to reduce them?
- Who should be in charge of developing the necessary standards?

A-1. STANDARDS

Good progress is being made on network standards. Our only concern in that regard is that the network conform to the principle standard (OSI model). It should allow the transmission of voice, data, facsimile (i.e., built-in error correction) and full motion video. It should also allow for sub-channels, principally for the simultaneous transmission of multiple information formats. Finally, it should easily interface at the local level with the most prevalent Local Area Network (LAN) standards. X-Windows client server model would be a good operating environment.

As far as data formatting and database software are concerned, we do not advocate presently any standards in this area. We view this entire area still as maturing and we feel that advocating standards at this time would only stifle innovation. There are however, some suggestions we could make in this regard:

Database Software -- Develop software so that the retrieval engine is functionally separate from the user interface, and so that the communication between the two is carried on by some type of standard command set -- perhaps a modified version of the Standard Query Language (SQL). Finally, it would be worthwhile to define a minimum set of these search commands that every search engine would understand. In this way, plenty of room for innovation would be left per individual developers, while the ability to communicate with the databases exists.

Data Formatting -- Data should be represented in standard form wherever possible.

- Text -- Standard ANSI or ASCII format.
- Compound Document Format -- TIFF and CCITT Gp IV are both standard ways to deal with multiple format page data. TIFF allows the user to define a page composed of many parts in differing formats.
- Color Images, Animation, Sound, or Full-Text Video -- These 'rich' data formats can be represented by proposed standard formats such as CD-ROM/XA, DVI or CD-I.

Standards adopted too early can create undesirable side effects by stifling creativity and erecting barriers to change. All standards should move through one of the existing standards-setting organizations. NISO and eventually ISO should be the ultimate standard setting/ratification bodies.

Q-2. *COMMERCIAL INFORMATION VENDORS*

Your company is one of many selling electronic data. In addition to University Microfilms, there are many online database companies, Dialog, Lexis-Nexis, and so on.

- *What kind of data products do you expect to be offered in the coming years: What is on the drawing board at UMI?*
- *Am I right in assuming that UMI would like to have access to the network we are discussing?*
- *How much would UMI be willing to pay for such access?*

A-2. COMMERCIAL INFORMATION VENDORS

UMI, in cooperation with publishers, anticipates that it will have nearly two thousand periodicals available as full-text image products by the end of 1990. The literature will be business, general, and scientific and technical, and will be tied into searchable abstract and index databases.

UMI would indeed look forward to accessing the proposed network, based on an appropriate fee. The fee could be transactionally based, or be computed on access time, or be a fixed fee. What works best can be worked out later. It is impossible for me to quote what UMI would pay for such access, because we do not know exactly what the economics for the network are. For example, we need to be able to model how much of various kinds of data can be simultaneously accessed by users before we could calculate what access to the network would be worth. Access costs would have to be competitive with other means of delivery, of course.

Q-3. ELECTRONIC COPYRIGHT

One big challenge posed by having books and journals online is copyright protection. How can an author know that his copyright will be protected if his work is online? How can he collect his royalties if anyone can print out his work with a push of a button? This could make the problem of illegal xeroxing look trivial. How can an author make sure someone does not plagiarize or alter his work.

- *You have faced many of these problems at your company. How have you taken care of the concerns of authors and publishers?*

A-3. ELECTRONIC COPYRIGHT

UMI contracts directly with copyright holders for permissions to publish or republish their information. We pay royalty fees to the copyright holders for the use of their materials. In the case of the electronic full-text products, we capture at the workstation level an abbreviated standard serial code for each article printed. This code is a mechanism to pay royalties to copyright holders and to evaluate what and how information is being used. In many cases libraries attach debit card boxes to the workstations and the patron pays for each page printed. Debit card readers or coin boxes are attached to most xerographic copiers in libraries today. The International Standard Serial Number (ISSN) for each serial volume and issue is also applicable to electronic networking.

Currently, UMI provides only "bit mapped" facsimile images of full-text pages, so that the text itself is not alterable electronically. This means that "what you see is what you get," to quote Flip Wilson. ASCII downloaded data is manipulable, however, and copyright holders have a legitimate concern that their information could be stolen, transmitted, or altered illegally. There is no immediate safeguard against the unlawful downloading and use of ASCII data. In some instances, online services permit downloading and charge a transactional fee, just as if it was printed. In addition some CD-ROM products permit downloading of bibliographic data on local workstations, as part of the subscription price.

Q-4. *NETWORKING AND DATABASES IN OTHER COUNTRIES*

France has made considerable investment in the development and implementation of a national online system known as Minitel, which reportedly allows 6 million French telephone users to easily access over 10,000 data bases.

- *Have its potential strengths or weaknesses been evaluated? What do you think of Minitel?*
- *What can the U.S. learn from the French that would help us to tackle our evolving systems?*

A-4. NETWORKING AND DATABASES IN OTHER COUNTRIES

MINITEL has been a showcase for videotex, but it is not clear that it has been a commercial success. The various videotex experiments here in the United States have been disappointing and investors have lost significant amounts of money. For more simplistic applications, the MINITEL terminal is perhaps acceptable, but it is relatively slow (only 1200 b/s, I believe). Intelligence and memory is quite limited. Many of the available databases are casual in nature. My understanding is that while the overall volume of usage is growing, the individual usage is flat, or even declining.

One lesson to be learned from MINITEL is the need for compatibility and standardization. For example, the Canadian and French videotex terminals are incompatible, which means higher costs for each. The second lesson is that people will buy information electronically only if it meets a real need.

Q-5. *NETWORKS VERSUS OPTICAL DISKS*

Your company makes optical disks which can be used by a single computer or accessed over a computer network. At some universities, your customers are starting to use your product on a central computer which is connected to a network and can serve users all over campus. In the future, a single optical disk library could serve users nationwide.

- *What are the benefits of having a centrally-managed online database versus having every user with his or her own set of optical disks? What are the disadvantages?*
- *What is the appropriate technology mix, and what applications are appropriate for the optical disk technology?*
- *When you look at the way that technology is moving, do you think things will become more or less centralized?*

A-5. NETWORKS VERSUS OPTICAL DISKS

We believe that there will be a mix between locally distributed databases and centralized services. Basically the mix will be a function of demand and economics. Heavily and widely used data will be required locally, whereas secondary and retrospective data could be accessed remotely. We have found with CD-ROM databases that user demand grew enormously. A single centralized service most likely could not accommodate economically millions of users simultaneously.

Information will not necessarily be cheaper just because it is delivered electronically. A printed page remains an inexpensive distribution medium, and it is cheaper to deliver a book on "dinosaurs" by overnight express, than it will be to send and print it electronically.

Senator GORE. Thank you very much.

Mr. Ted Nelson is from Autodesk, Incorporated, one of the inventors of the Hypertext system for organizing, combining and transmitting video images, text, and other electronic data. We appreciate your presentation here, Mr. Nelson.

STATEMENT OF TED NELSON, AUTODESK, INC., SAUSALITO, CA

Mr. NELSON. Thank you. I am not only honored to be here, I am very moved because for so many years I have been consigned to the lunatic fringe of computerdom—in fact, thought by some to define the lunatic fringe of computerdom.

Senator GORE. And you feel at home here today?

Mr. NELSON. Well, when I said personal computers would replace the typewriter, they thought I was crazy and when I said Hypertext is the new step in literature, they said I was crazy. And when I said on-line libraries were coming, they said I was crazy. And when I said open Hypertext publishing was coming, they said I was crazy. And now they call me a visionary. So, that is fine.

There are two cures for paranoia which is believing what others do not believe. The first cure, of course, is to change your mind and start believing what the other people believe. And the much better route is to persuade them, which seems to be happening.

The Hypertext concept which has caught on very widely is simply that nonsequential writing is going to be a very important new medium and hypermedia, another current term I coined long ago, which means interactive presentations—you were speaking yourself, Senator, of the kid being able to explore dinosaurs and go from object to object like a frog going from lily pad to lily pad looking at this and then at that, studying one thing and then another.

I think this is certainly the way we are going to go now that we have the beginnings of technology in place. But having it on a video disk is only a—you see the good news with video disks is you have 400 million bytes—40 million bytes on your desk top. And the bad news is that you only have 40 million bytes on your desk top.

Senator GORE. Which translates to about an hour's worth of videotape.

Mr. NELSON. If you are running it sequentially.

Senator GORE. If you slice it up in little still pictures, then it is millions.

Mr. NELSON. Well, hypermedia has no running time. In other words, you can spend your whole life contemplating one sentence, of course, or you can hop around like crazy.

But the real issue is what are we going to do when we have the capability and what I mean by open Hypertext publishing means this: that not only are individual authors creating materials in which material connects every which way and the reader can follow it, not only are they creating it so that the sophisticated reader can best explore this material. Imagine, for example, if the Congressional Record being rather than one long sequence of talk and patter were organized——

Senator GORE. I beg your pardon?

Mr. NELSON. Scented talcum powder or organized into sections which you could branch to in different ways and find the parts that you wanted threaded through in many different ways.

Okay, this is different from the data base notion which we have heard a lot about today which means searching through materials that you can riffle that are like virtual file cards. The hypertext notion is jumping from one thing to another, but open hypertext means that different people can start contributing all over.

Imagine, for example, one of the projects we are working on, a hypertext of World War II to which every participant is invited to make a contribution attaching it to that story which is most relevant and to the generals you met or the people you shot or whatever you did. The ability to create, to publish disagreements that are attached right to the article that you are disagreeing with, endorsements, reexplanations, taking the material that is already there and saying no, this is a better way to look at it so that rather than having the separate contributions in far-flung places throughout the corpus, they can actually be attached where people can find them.

The easiest way to say it is I have a dream, okay. The year 2020, the 2020 vision. A billion people are at their screens around the planet, each one able to bring forth any paragraph, any sentence, any fragment, any footnote, any illustration essentially in the length of time it takes you to reach the other side of the country on the telephone.

Now, it strikes me as very funny that people are not surprised that they can reach somebody on the other side of the continent by telephone and yet the idea that they could draw these things from a massive library seems very startling.

We are going to have to have this kind of controversy management system to prevent the ecological disasters of tomorrow and to deal with the decision processes of the future.

What my Project Xanadu has been concerned with for the last 29 years—I stress that I have been doing this for 29 years—has been trying to create the software for this world of the future beginning with, shall we say, a data management system which will be sold by Autodisk, by the way.

This was a gang of ne'er-do-wells operating in an atmosphere somewhere between Camelot and the Manhattan Project for a long time. We achieved legitimacy last year when Autodisk, a very rich company, bought us and has put several millon dollars into it.

What it is going to do is essentially change the way that data structures are used. Instead of the silly files which we use now on computers which have 8.3 characters in their name and everyone in the world who uses a computer has a drawerful of things that cannot be found, little disks on which things have incomprehensible names whose connections are unknown.

The problem is to manage this connection to learn how everything attaches to it together. And so, creating software for managing these connections and these attachments is what we have been about.

But the purpose of this software has not been merely to make things easier for computer users. The purpose of this software has been to create the open hypermedia publishing of tomorrow where the billions of people can get at the connected material and make their comments.

So, that brings me to the last point. In approximately two years—and, again, this is conjectural—we have a handout over here—we will be starting what we call public access Xanadu, which will be the publishing system.

We are not asking for government subsidies for this. We want to use your high band with network because we believe we can absorb—we believe that the capacities of the information providers of this nation are going to absorb your band within probably three months or so after it opens.

It is like the first Xerox machines. You may recall—I was in various committee meetings where they were trying to decide whether a university could use a whole Xerox machine, you see. This is true. Back in 1960 they said well, we only make three copies a month. Why would we ever do that?

So, within a week, of course, the entire capability was sopped up. Similarly for open hypertext publishing. Let me state that we are going to allow people to publish materials in an open hypertext format starting roughly in the middle of 1991. We will have franchise operation information stands rather like McDonald's called Xanadu stands where you can come in, start your account, and use this from home, either publishing or pulling in stuff from the base which is published already.

And we have been building this software, looking forward to the billions of people using billions of documents, and this required a very different software direction.

And we look on you to give us the pathways, the superhighways for information to carry this stuff. But we think that the private sector may be able to do a lot more than you think for the digital libraries of tomorrow.

Thank you.

[The statement follows:]

Statement to the Senate Commerce Committee by
Theodor Holm Nelson, Distinguished Fellow, Autodesk, Inc., and Director of Public-Access Xanadu, 15 September 1989.

I am grateful for this opportunity to represent my company and the aspirations of many people in the computer field. I think I speak for a large number of sophisticated computer people and others concerned with communication in the world of tomorrow.

TITANIC CHANGE

We stand at the threshold of a titanic change in the use and availability of the written word, a new era of electronic literature.

I am not talking about "data bases," as we already see them in commercial use, for those have reached roughly the limit of their usefulness. The kinds of searching, flipping and scanning that they make convenient are useful in many ways to researchers, both academic and commercial, but they are difficult and largely irrelevant to the main uses of the written word: learning, study, finding out different points of view, browsing and skimming.

Nor am I talking about "videotext," a collective term for various low-grade services that have been devised for people who are basically uninterested in the written word.

Nor am I talking about the alleged joys of "CD-ROM," the selling of computer disks with a lot of stuff on them for desktop use.

Nor am I talking about "electronic mail," the increasingly popular use of computers to push letters back and forth between users at high speed, permitting individuals to exchange many communiques a day.

Nor am I talking about "teleconferencing," which in its current form is like long electronic scrolls to which different people add comments.

Nor am I talking about "bulletin boards," computer storage systems used to hold and forward materials placed on them by random outsiders.

I am talking about **the coming hypertext revolution**.
Hypertext, I believe, is not just the latest fad in computerdom. It is the next step in civilization, and it represents a quantum step in the use, delivery and meaning of the written word comparable to that offered by Gutenberg's printing press.

The hypertext revolution will be that of on-line publishing to computer screens, by and for people using computer screens.

HYPERTEXT

Hypertext means non-sequential writing. It is practical only on computer screens.

Writing, as I believe Marshall McLuhan observed, derives its form from the medium of paper. Writing has been sequential because books were necessarily sequential. We have put numbers on our pages in the past because that was the only way to find your way around those physical objects of paper called books and magazines. But on the computer screen we no longer need numbered pages, since the reader at the screen can jump to whatever he or she wants to see next. And this leads to deep changes in the way we organize what is written and shown graphically.

The sophisticated reader does not usually read sequentially. Picking up a book, he or she flips through it, looks at the beginning, the end, the middle, the pictures; and with each step, learning more, decides what to do next. But, curiously, we still *write* as though the reader is going to read the first word, and then the second, and so on.

The computer screen is changing all that. Already, in numerous systems (such as Apple's Hypercard and Owl's Guide), users are reading and writing in nonlinear form.

THE UNIVERSAL REPOSITORY

It was in anticipation of these developments that I began this work twenty-nine years ago, and my work since then has been concerned with expanding the hypertext idea to a universal publishing system. Now the world is ready, and in the meantime the group I have gotten together has been preparing the software.

The best way to explain the overall vision is in terms of what we do not have yet.

There is not currently any convenient place where you can put data so that people can send for pieces of it and automatically pay you a royalty. Nor is there currently any place where you can find someone else's data and link your own data to parts of it. Nor is there currently any place where you can publish electronically a comment on something else that is already published electronically.

These things we intend to create, in the form of a unified repository for the storage and transmission of published data, with automatic proportiaonl royalty payments to the publisher.

We hear of people creating special data repositories for music, for literature, for history, for scientific data, for graphics. This is crazy. Everything ties together. The only reasonable approach is to create a unifying repository where ALL types of data may be stored. And the only reasonable person to keep track of the integrity of each piece of data is its owner or publisher, who must also be responsible for its content (just as the publisher is now).

Such a repository may grow, we think, to hold all collectively all the writings and storage of humankind.

Such a repository cannot be kept on a big computer mainframe, because it is going to be immense, and because it has to be kept safe. It will be too big to put in one place and it would not be safe if it were in only one place. So it must be distributed on a network of many computers.

But a new kind of software is needed for this plan: a computer program that can keep track of any and all connections among these data, and keep these connections orderly as this great network of stored material grows and grows and grows.

PROJECT XANADU

For years I have called this work "Project Xanadu," and it has consisted of both aspects: the plan for a world-wide respository, and the software necessary to make it happen. Over the years the plan became more and more detailed, and I managed to assemble an extraordinary team to do the work-- the best and the brightest and most idealistic programmers I could find. Notable among these are Roger Gregory, Mark Miller, Roland King, and Eric Hill: I brought them together a decade ago and paid them nothing, and they brilliantly designed and programmed an overall system with extraordinary possibilities for immense worldwide growth. More recently they were joined by Erik Drexler, the author of *Engines of Creation*, and together they have brilliantly designed the present system. All this is described in my book *Literary Machines*.

We divided Project Xanadu into two parts: Xanadu Operating Company, Inc., to create the software and market it to business and industry, and what is tentatively called PAX, Public-Access Xanadu, to market this repository publishing scheme with automatic royalty.

Last year, after twenty-eight years of work, Xanadu Operating Company was bought by Autodesk, Inc., a world leader in software development. The chief product of Autodesk is Autocad®, the world-wide standard for computer-aided design, which holds about 50% of the world market. Not merely throughout the United States, Europe and capitalist Asia, but even in the Soviet Union and China, the Autocad program is the standard.

The purchase of Xanadu Operating Company by Autodesk has meant new resources for the completion of the Xanadu program and its sale to business and industry, which will begin in 1990. This signals a new era in the interconnection of data. No longer must large projects and complex information be divided into separate data chunks, called "files," which must be stored under a lot of different names and be kept track of on paper. With Xanadu™ storage it will be possible to bind together huge conglomerates of data, containing many different types of data format, into unified structures with alternative versions, historical backtrack through changes, and with variants and overlays and different viewpoints on the same material. But most important will be its *connections:* inks among parts of the material and links to outside structures. Users may take note of relationships, make comments, and follow these connections interactively on their computer screens. (Through another kind of connection, called *transclusion*, the same materials may be used in many different structures without copying those materials in the new places.)

I would like use this opportunity to explain and advertise this system further, but that would be irrelevant and improper. I am here to explain how this system foretells a vast growth in the transmission of digital information, and a vast growth in the need for bandwidth throughout the United States of America.

CONNECTED INSTANT PUBLISHING

There are already numerous hypertext products on the market, including GUIDE from OWL International, Hypercard from Apple, Notecards, Hyperties, and others. All these suffer from mutual incompatibility. This means that people creating works in one hypertext system cannot connect them to works in others. This is only one aspect of the brutal incompatibilities we now see in the computer world.

Universal interconnection and possible standardization.

The Xanadu program should be able to change all that, by providing linkage forms (and new methods of data standardization) that will allow materials with many origins to share storage, to be linked together, and to be used together. Thus there is some hope for a new era of compatibility.

Open hypertext publishing.

More than this: we may look forward to *open hypertext* *publishing*, where an author or commentator may add links to material prepared by someone else. This is a unique and potentially sweeping new aspect of publishing. I have discussed its ramifications, and ways to keep it orderly, in my book *Literary Machines* and elsewhere.

We live in an age of onrushing change, with greater problems than ever before; to some of us it appears that only this new prospect of open hypertext publishing, with its potential for clarifying and sifting information in a great worldwide round-table forum, offers us hope of survival.

Instantaneous delivery.

Startling to most people is the notion that published materials can be delivered right away-- potentially in seconds-- to a vast network of users, on their demand. But the technology exists; it is merely the problem of designing an integrated and useful service that holds this back. We believe that the type of linked and transclusive hypertext publishing offered by the Xanadu system is the integrated and useful service that the world has been waiting for.

Universal access.

Materials stored on the PAX™ publishing system will be available throughout the world´s telephone networks, as well as over specialized connection systems that it may be convenient to set up. So essentially anybody can get at and use these materials.

In principle this is no different from your ability now to order books, journals and reprints from various sources-- publishers, libraries and retrieval services-- throughout the world. But the rapidity and ease of access should make a dramatic difference in the way that scholarly, scientific, artistic and other material will be accessible to everyone.

A franchising model.

We expect that the PAX system will be expanded, not through the backing of governments or libraries, which are very short on funds, or through large corporations, which are very short on innovation. We expect this to grow through a well-known and popular system known as franchising-- the system that has made the golden arches of McDonald´s better known than the Arch of Triumph or the great arch of St. Louis. Again, I will not dwell on this marketing point, which I have made in my writings.

THE MANIFEST DESTINY OF LITERATURE

Fans of the Xanadu project, those who believe in it, consider it to be not just an off-the-wall technical development, but something very different: the natural next stage of the written word. Hypertext is not an exotic idea out of Silicon Valley (as some have alleged), but a form of writing that has been implicit in the written word since it began. But it could not unfold until a technical basis could be found for it, a technical basis we have now found in computer storage and on the computer screen. And the instantaneous delivery of literature-- literature in the broadest sense, the dissemination and preservation of prepared information packages that can include graphics, sound, video, statistics, laboratory information and anything else we ever digitize-- should seem no more exotic to us than the instantaneous delivery of the human voice across the telephone, or the instantaneous delivery of the human comedy by television.

THE NEED FOR BANDWIDTH

This brings me to the punch line. The Xanadu system, or something very like it, is inevitable and vital to help us with information to conquer the problems of tomorrow. But if millions of people are going to use this new form of publishing-- and I am certain that they are, both as receivers and creators-- then we are all going to need all the bandwidth we can get. Fiber is good for us, both medically and electronically. I say let´s put in all the information piping we can get. As a scientific community we need it. As a free people we need it. From every mountainside let circuits ring, and let the chips install where they may.

Senator **Gore**. Let me just say that I have, I hope, a full appreciation for what the private sector can and will do in this, and I have said repeatedly that I think this entire project ought to be in large part transitioned to the private sector.

When you get over the hump and you have the capacity for user fees and you have the ease of continuing self-financing, I fully agree with that. I think there are a lot of functions that will probably have to remain with the government, such as standardization and copyright issues. But I agree with you on the dynamic role of the private sector in all this.

I am really excited by the presentations on this panel and those that have been made throughout the day.

I always finish one of these hearings realizing that we could have had three times as much time allotted and still not gotten all the useful interchanges accomplished. So, I do reiterate my apology for not having as much time as we would like to spend with this panel. But they were very provocative presentations.

Senator Robb and I both had planned on the hearing finishing up like about 45 minutes ago, so we are both late to other things. But I would like to thank all of our witnesses today and announce that we are going to move forward rapidly with the legislation as rapidly as we possible can.

The legislation has been introduced in the exact same form by the chairman of our counterpart subcommittee in the House of Representatives.

The Bush administration has put its stamp of approval on this initiative, so we are going to move forward very rapidly. We will be drawing on the continuing advice of the people here on this panel and the other panels for help in that.

Before we close, did you have any comments?

Senator **Robb**. Mr. Chairman, I echo your comments, and I appreciate the fact that you held this hearing, and I join you in thanking this panel and the others for adding to our knowledge. And I look forward to working with you.

Senator **Gore**. Thank you very much. The hearing is adjourned.

[Whereupon, at 12:45 p.m., the subcommittee adjourned, subject to the call of the chair.]

ADDITIONAL ARTICLES, LETTERS, AND STATEMENTS

STATEMENT OF HON. ROBERT KERREY, U.S. SENATOR FROM NEBRASKA

Mr. Chairman and members of the Subcommittee I am pleased to support S. 1067, The National High-Performance Computing Technology Act of 1989. This forward-looking initiative will help create the necessary infrastructure and network to revolutionize the way we live our daily lives.

The realization of this bill's goals are important not only for reasons of economic competitiveness or scientific research, but perhaps more importantly for the educational opportunities it will provide. An extensive computer network that provides access to databases -- text, video, sound and others -- allowing ordinary citizens to tap into the vast spectrum of information will mark a qualitative leap in the availability of information.

I appreciate Senator Gore's desire to develop computer technology that is not driven by military or strategic thinking. All too often our technological initiatives are oriented to achieving military advantage. Peaceful applications of technology have fewer advocates than military applications. A tool which will permit a 10-year-old to explore and achieve greater understanding with more complete freedom may appear less exciting than the roar of a fighter bombers or the killing capacity of a new weapon.

Fear of a perceived enemy creates an urgency that requires the commitment of vast resources to the cause. The desire for new knowledge is a more passive urge with less political value. Moreover, new discoveries very often threaten the status quo. There are few things worse than having your children come home from school after learning something that has caused them to conclude their parents are wrong.

While this initiative is important, it should be viewed as a first step that ultimately leads to the achievement of our ultimate endeavor: creating a system that serves children. Children both in urban and rural areas who have traditionally found themselves dependent on limited resources all of a sudden can have access to virtually unlimited material. Materials that will allow them to learn a language, to explore science, to improve mathematical skills, or to pursue any other line of questioning.

The objective of any system we develop must focus on the needs of a 10-year-old rather than on the economic needs of businesses. I believe that it will work for business if it works for the child. I do not believe that the opposite will necessarily hold.

I look forward to working with Senator Gore and other members of the Subcommittee to see that the bill is enacted.

Association of American Publishers, Inc.

1718 Connecticut Avenue, N.W
Washington, D.C. 20009-1148
Telephone 202 232-3335
FAX 202 745-0594

STATEMENT FOR THE RECORD

FROM THE

ASSOCIATION OF AMERICAN PUBLISHERS

FOR THE SEPTEMBER 15, 1989 HEARING

ON

COMPUTER TECHNOLOGY

AND

THE NEW AGE OF INFORMATION

There are several concerns held by book and journal publishers as they consider the new technologies represented by a National Digital Library System. These concerns are identified below:

SYSTEM INTEGRITY:

Before a publisher will trust data to such a system, there must be assurances about the integrity of the system. A work must be able to be maintained in content exactly as it was when entered unless the copyright holder authorizes changes.

The pertinent provision of the copyright law stipulates that no "derivative work" (work derived from the original) can be created without permission of the copyright owner. Hence, the unit of information placed in such a system must be secure from any impermissable tampering while it is in the system. However, the creator of the document (author or publisher) should be able to take the original work out of the system and modify it to create an amended amended version, and should be able to allow changes to be made by those who are authorized to do so by license or other arrangement with the copyright holder.

ATTRIBUTION:

Along with the issue of integrity is the question of attribution. Will excerpts of the information stored in the system in machine-readable form be able to be viewed or stored locally (i.e., on a reader's personal terminal)? If so, will the reader be able to store excerpts of the work or only the entire unit of information? If excerpts are permitted, could they be dissociated from data on attribution so that at a later reading, the user will have no way to track the source of the excerpt (in other words, the copyright holder)? The system should provide that such attribution remains inseparable from the article or any part thereof.

LOSS OF FREE WILL:

Could a publisher be co-opted by establishment of a national digital library system that is ostensibly voluntary, but in point of fact is not? In other words, if an electronic system becomes the prime or only place people check for bibliographic references, does a publisher's refusal to participate effectively shut the non-participant out of the marketplace? Such a system could coerce publishers into accepting terms that would otherwise be unacceptable and possibly harmful to the continued existence of the underlying materials.

STANDARDS:

In order to create a National Digital Library, there must be standards for storage of information, for search protocols, for display and presentation. For example, standards concerning treatment of graphics raise the question of how the standards will be set and whether they will be equally appropriate for a highly technical document that needs optimum resolution for graphics and for less graphically oriented materials? Is there a way to vary the graphics to meet the needs of the publisher without driving costs up?

REVENUE AND LOSS POTENTIAL:

In order to encourage development of a National Digital Library System, there must be incentives to encourage authors and publishers to put documents into the system. Publishers of traditional journals in print form must be convinced that the revenue potential from use of their materials in the system will make up for lost subscription income from the paper copies no longer needed.

Similarly, if a publisher decides not to participate, does the fact that his competitors choose to participate coerce participation at the threat of lost market share? What marketplace pressures are brought to bear? What existing markets could a National Digital Library System displace?

PARTICIPANTS:

Who is best suited in the new electronic media marketplace? If private sector entities alone are involved, then development is based on the degree of investment the private sector is willing to make. Costs would have to be structured to allow investors to recoup investment. However, if the government develops the system with tax money invested in research and development and the technology is available to all private sector entities, will the costs be more reasonable and will the resulting larger market mean more income for participating publishers?

DESIGN AND STRUCTURE MUST ANTICIPATE USE OF COPYRIGHTED MATERIALS:

S. 1067 contemplates the creation of a National Research and Education Network. Whether the initial information stored in the system is public domain government data or copyrighted materials does not change the fact that a system that ultimately intends to include copyrighted material must consider copyright issues at its inception.

Use of public domain materials only could create an expectation in the user community that any information in the system could be used freely without regard to copyright. This expectation would be hard to reverse when the system progresses to the inclusion of copyrighted materials.

OTHER ISSUES:

Other intellectual property concerns have direct parallels in the print world. Laws exist to deal with an unscrupulous colleague, who in an attempt to discredit another, issues false information in someone else's name. A National Digital LIbrary System will not have to track usage to be able to confirm this. All that the plaintiff would have to prove is that the defendant had access to the information in the system. If someone manipulates an existing work and creates a new work, existing copyright law applies - and the new "author" would be judged by the substantial similarity and access tests already in existence. Public policy and privacy questions should be considered, but these can be viewed from the print perspective.

CONCLUSION:

The Association of American Publishers submits this statement for the record in order to create an awareness of the issues outlined above. AAP takes no position on the National Digital Library System. However, if such a system is to be created, and if it is to include copyrighted works, then any inquiry into its development must include concern for the rights of the copyright holders of the works that will be in the system.